

Sparse Shape Reconstruction

Alireza Aghasi and Justin Romberg*

Abstract

This paper introduces a new shape-based image reconstruction technique applicable to a large class of imaging problems formulated in a variational sense. Given a collection of shape priors (a shape dictionary), we define our problem as choosing the right elements and geometrically composing them through basic set operations to characterize desired regions in the image. This combinatorial problem can be relaxed and then solved using classical descent methods. The main component of this relaxation is forming certain compactly supported functions which we call “knolls”, and reformulating the shape representation as a basis expansion in terms of such functions. To select suitable elements of the dictionary, our problem ultimately reduces to solving a nonlinear program with sparsity constraints. We provide a new sparse nonlinear reconstruction technique to approach this problem. The performance of proposed technique is demonstrated with some standard imaging problems including image segmentation, X-ray tomography and diffusive tomography.

Keywords: Sparse Shapes, Parametric Level Set, Nonlinear Compressed Sensing

1 Introduction

In many imaging applications, the main objective is to identify and characterize regions of interest in a given domain. This characterization is usually based on some property defined over the imaging domain. For instance in image segmentation [40, 42], this property is directly or statistically related to the pixel values and the partitioning is usually meant to aggregate similar regions. In shape-based inverse problems [18, 24, 32], the property of interest corresponds to a spatial physical parameter which needs to be determined by some indirect observations. Here a shape-based characterization delineates inclusions and obstacles causing a contrast in the values of the spatial parameter. In the basic binary case the shape characterization problem is formulated as partitioning a compact imaging domain $D \in \mathbb{R}^q$, into $\tilde{\Omega}$ and $D \setminus \tilde{\Omega}$, where $\tilde{\Omega} \in \mathbb{R}^q$ is a closed set representing the object(s) of interest.

*School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA. Emails: aaghasi@ece.gatech.edu and jrom@ece.gatech.edu.

In this paper we propose a new approach to shape-based modeling. The idea is to characterize the target inclusion (or partition) as a composition of given shape prototypes. These prototypes may either be shapes with simple geometries or shape priors that are likely to be related to the structure of the target geometry. The idea of composing basic shapes to form more complex structures is quite intuitive. The challenge, is to formulate shape composition concretely and then incorporate it into the reconstruction process in a computationally tractable manner.

For the purpose of shape composition, we consider a mechanism to merge shapes and exclude undesirable regions from the aggregate to generate $\tilde{\Omega}$. More formally, given a set of prototype shapes, $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_n$, which are known closed sets in \mathbb{R}^q , we formulate the shape composition process as applying set union and relative complement among a selected number of shapes in the reference set to generate a region that approximates $\tilde{\Omega}$. A brute force approach would entail exploring all possible selections and composition possibilities to find a good fit. To have an idea about the accuracy of an estimate of $\tilde{\Omega}$, we specifically focus on variational shape-based problems where a score (value of an energy functional) is assigned to every closed region $\Omega \in D$ and $\tilde{\Omega}$ is meant to minimize such functional.

This type of brute force search is in general computationally intractable. In this paper we will propose a relaxation formed by defining so called *knoll* functions which characterize $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_n$ by their support. We will show that for a smooth energy functional associated with the shape-based problem, this relaxation can convert the problem into minimizing a smooth nonlinear function in \mathbb{R}^n .

To encourage our target shape to be as simple as possible, we impose sparsity constraints on the resulting nonlinear cost. Accordingly, we propose a minimization technique inspired by the idea developed by van den Berg and Friedlander in [47]. The original idea in [47] developed for linear inverse problems is merely applicable to the corresponding quadratic least squares cost. What we will put forth is a generalization of this approach applied to quadratic estimates of the nonlinear problem at iterative stages of approaching a minima.

The organization of this paper is as follows. In the remainder of the introduction section we provide a rather extensive overview of variational shape-based methods, specially in context of active contours, pixel based and parametric level set techniques. After providing this background, in Section 2 we present the shape composition idea and discuss employing a so called *pseudo-logical* shape interaction property which will assist us in developing a relaxation to the corresponding combinatorial problem. Section 3 is devoted to developing a Gauss-Newton type sparsity promoting algorithm to solve the resulting nonlinear problem. Finally, in Section 4 we consider applying the proposed technique to some image segmentation and shape-based inverse problems¹. The image

¹The Matlab code for the proposed algorithm is available at <http://users.ece.gatech.edu/>

processing applications considered are segmentation with missing pixels and machine text recognition. We also consider an example of medical X-ray computed tomography and an archaeologic resistance tomography problem, both with limited available data, where majority of the state of the art techniques would not be able to provide satisfying reconstructions.

1.1 Background on Variational Shape Reconstruction

In the past decades considerable effort has been devoted to exploring variational techniques applicable to the shape-based characterization problem [15, 16, 23, 31, 36, 46]. The main intuition behind a variational approach is forming an energy functional, minimizing which tends to solve the characterization problem. More specifically, for an arbitrary bounded set $\Omega \in D$,

$$\tilde{\Omega} = \arg \min_{\Omega} \mathcal{E}(\Omega), \quad (1)$$

where \mathcal{E} is the underlying energy or cost functional. An example of \mathcal{E} in the context of inverse problems is the data-model mismatch functional

$$\mathcal{E}(\Omega) = \|v - \mathcal{M}(\Omega)\|_{\mathbb{S}_v}, \quad (2)$$

where v represents the observed data, \mathcal{M} denotes the model mapping the shape geometry to the observations and \mathbb{S}_v is a Hilbert space associated with the data. One of the main motivations in representing problems in variational forms is the chance of using descent optimization techniques.

Of course a straight-forward approach in determining a region Ω is determining its boundary \mathcal{C} . An early technique of this type is the active contour model (snakes), where the shape determination amounts to determining the parameters associated with a spline representation of \mathcal{C} [31]. Starting with an initial shape representation, this process takes steps along the descent direction of the energy to evolve the shape towards an optimum state. More specifically, by defining an artificial time t , an initial contour $\mathcal{C}(0)$ is evolved in time and according to

$$\frac{\partial \mathcal{C}(t)}{\partial t} = - \mathcal{E}'(\mathcal{C})[\mathcal{C}(t)], \quad (3)$$

to find the boundary of the shape that locally minimizes \mathcal{E} [14]. Here $\mathcal{E}'(\mathcal{C})[.]$ is a linear operator representing the Gâteaux derivative (or first order variations) of the energy functional with respect to \mathcal{C} .

Another well known shape-based technique is the level set method [37, 39]. The main advantage of level sets over earlier methods such as snakes is their topological flexibility, dispelling the need to any prior assumptions about the number of connected components

in $\tilde{\Omega}$. Here the zero level set of a Lipschitz continuous function, ϕ , is used to identify \mathcal{C} . More specifically the objective is to determine ϕ such that

$$\begin{cases} \phi(x) \geq 0 & x \in \tilde{\Omega} \\ \phi(x) < 0 & x \in D \setminus \tilde{\Omega} \end{cases} . \quad (4)$$

By using a map as (4) our geometric problem is cast as the calculus of variations problem

$$\tilde{\phi} = \arg \min_{\phi \in \Theta} \mathcal{E}(\Omega_{\phi}), \quad (5)$$

where Ω_{ϕ} is the shape resulted at the zero level set of ϕ , and Θ represents a certain function space to prevent ill-conditioning. Most level set implementations consider elements of Θ to be signed distance functions (SDFs) [37].

For the level set methods, minimization of (5) is performed by evolving an initial level set function, $\phi(x, t) = \phi_0(x)$, through the Hamilton-Jacobi equation

$$\frac{\partial \phi(x, t)}{\partial t} + V(x, t) \cdot \nabla \phi(x, t) = 0. \quad (6)$$

This equation results from a straightforward differentiation of the front equation $\phi(x, t) = 0$ with respect to t . In this equation $V(x, t) = dx/dt$ is a speed function applied to the zero level set of $\phi(x, t)$ and taken in the descent direction of \mathcal{E} to reduce it as ϕ evolves. At every iteration, through a proper speed function extension or re-initialization, the level set function is assured to remain an SDF. We refer the reader to [10, 23, 37] for more details about implementation of this technique.

Level set methods are among the most successful techniques in shape representations, mainly due to their topological flexibility. They perform remarkable for a large class of image processing applications [20, 38], however, in general there are implementation concerns and performance deficiencies associated with them. Numerically speaking, evolving the level set function requires discretizing it over a dense grid of pixels and updating the pixel values over discrete time frames (iterations) [37]. Using pixels to parameterize the level set function brings a large dimensionality to the problem which may severely affect the overall performance of the method in dealing with ill-posed functionals and inverse problems (see examples in [1]). Although in principle level sets are topologically flexible, it is usually not possible to create holes or shapes away from the boundaries of the evolving shape [9]. Moreover, implementation complexities such as speed function extension and maintaining ϕ as an SDF are usually the inevitable components of this technique.

To overcome these problems and still take advantage of the topological flexibility, researchers have started considering parametric forms for the level set function [1, 5]. This strategy not only reduces problem's dimensionality by making it parametric, also provides a predetermined functional form to dispel the need to re-initialization. An

example of such effort is the work by Bernard *et al.* [5], where according to (5), Θ is the space spanned by pre-assigned B-spline functions. The problem in this case usually amounts to the energy minimization

$$\tilde{\alpha} = \arg \min_{\alpha} \mathcal{E}(\Omega_{\phi(x, \alpha)}), \quad (7)$$

where α is the vector of parameters associated with the parametric level set function $\phi(x, \alpha)$.

A main advantage of parameterizing the problem as (7) is casting the shape characterization problem as a classic finite dimensional minimization problem, where gradient descent or even Newton type methods may be applied. For sufficiently smooth functionals, gradient and Hessian of \mathcal{E} with respect to elements of α may be conveniently calculated by following the chain rule in Fréchet spaces [1]. More specifically

$$\frac{\partial \mathcal{E}}{\partial \alpha_i} = \mathcal{E}'(\phi) \left[\frac{\partial \phi}{\partial \alpha_i} \right], \quad (8)$$

and

$$\frac{\partial^2 \mathcal{E}}{\partial \alpha_i \partial \alpha_j} = \mathcal{E}''(\phi) \left[\frac{\partial \phi}{\partial \alpha_i}, \frac{\partial \phi}{\partial \alpha_j} \right] + \mathcal{E}'(\phi) \left[\frac{\partial^2 \phi}{\partial \alpha_i \partial \alpha_j} \right], \quad (9)$$

where the linear and bilinear operators $\mathcal{E}'(\phi)[\cdot]$ and $\mathcal{E}''(\phi)[\cdot, \cdot]$ are respectively the first and second Gâteaux derivatives of \mathcal{E} with respect to ϕ . We refer the reader to [1] for more details on a parametric level set representation.

Although a parametric form solves many concerns with traditional level sets, the problem still remains on choosing a suitable parametric form (e.g., a suitable set of basis functions) and efficiently determining the number of terms. Clearly a dense basis set to increase the resolution of the reconstructions may bring redundancy and ill-conditioning to the problem as the case with pixel based level sets.

In the sequel, we propose a new way of approaching the shape problem. We extend one of the basic ideas developed by Aghasi *et al.* in [1] and provide a new shape reconstruction technique applicable to a large class of problems. The method ultimately finds a parametric level set form, however, we link it to a systematic algorithm of choosing the right parameters in the course of reconstruction.

2 Basic Shape Composition Idea

In solving (1) for the optimum shape Ω , an intuitive idea would be to consider a collection of fixed shapes and reconstruct Ω by applying basic set operations on them. More specifically, consider the shape collection $\mathcal{D} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{n_d}\}$, where for $i = 1, 2, \dots, n_d$, elements \mathcal{S}_i represent closed known regions (shapes) in \mathcal{D} . We name \mathcal{D} as the *shape*

dictionary or simply the dictionary. Suppose through an oracle we know what the true shape Ω is. To logically (in the sense of set operations) express Ω in terms of elements of \mathfrak{D} , one way of approaching the problem is to select an appropriate set of shapes indexed by $\mathcal{I}_\oplus \subseteq \{1, 2, \dots, n_d\}$ and apply the union operation over them to form the more bulky superset $\bigcup_{i \in \mathcal{I}_\oplus} \mathcal{S}_i$ for Ω . We now start to choose shapes indexed by $\mathcal{I}_\ominus \subseteq \{1, 2, \dots, n_d\}$ to carve out portions of $\bigcup_{i \in \mathcal{I}_\oplus} \mathcal{S}_i$ and make it a better approximation to Ω (in the sense of reducing \mathcal{E}). To be concise, for a given dictionary \mathfrak{D} we define the objective as the combinatorial problem of searching among elements Ω of the form

$$\Omega_{\mathcal{I}_\oplus, \mathcal{I}_\ominus} \triangleq \left(\bigcup_{i \in \mathcal{I}_\oplus} \mathcal{S}_i \right) \setminus \left(\bigcup_{j \in \mathcal{I}_\ominus} \mathcal{S}_j \right), \quad (10)$$

and find the suitable index sets \mathcal{I}_\oplus and \mathcal{I}_\ominus that minimize $\mathcal{E}(\Omega)$.

This simple idea is inspired by approximation theory of suitably expressing a function as a linear combination of some given basis functions. Here instead of adding and subtracting the basis terms with suitable weights, suitable elements of \mathfrak{D} are combined through the union and relative complement to provide the shape approximation.

We would like to highlight that more complex reference forms other than (10) may be considered. We however maintain simplicity by suggesting this form and assuming that \mathfrak{D} is rich enough that (10) still provides the desired flexibility in shape representation. For instance if the dictionary is poor, consisting of only two shapes \mathcal{S}_1 and \mathcal{S}_2 , and the true shape is $\tilde{\Omega} = \mathcal{S}_1 \cap \mathcal{S}_2$, none of the possibilities in the form of (10) would be able to express $\tilde{\Omega}$. However if $\tilde{\Omega}$ is already among the elements of the dictionary or we consider a richer dictionary consisting of \mathcal{S}_1 , \mathcal{S}_2 , $\mathcal{S}_3 = \mathcal{S}_1 \setminus \mathcal{S}_2$ and $\mathcal{S}_4 = \mathcal{S}_2 \setminus \mathcal{S}_1$, we can find $\tilde{\Omega}$ by exploring different possibilities of the form (10) since $\Omega = (\mathcal{S}_1 \cup \mathcal{S}_2) \setminus (\mathcal{S}_3 \cup \mathcal{S}_4)$.

Selecting appropriate elements for \mathfrak{D} may be based on the level of prior information about the geometric features of Ω . The shapes \mathcal{S}_i may be simple geometries that combine to form a more complex structure (e.g. see Fig 1(a)); a large collection of shape possibilities among which the true shape needs to be determined (Fig 1(b)); or a combination of both cases.

To bring this idea into application we need to develop a computationally tractable algorithm that performs the search among possible shape combinations and minimizes the functional in (1) using a sufficiently sparse set of elements in the dictionary. More formally we reformulate (1) as

$$\{\mathcal{I}_\oplus, \mathcal{I}_\ominus\} = \arg \min_{|\mathcal{I}_\oplus| + |\mathcal{I}_\ominus| \leq s} \mathcal{E}(\Omega_{\mathcal{I}_\oplus, \mathcal{I}_\ominus}), \quad (11)$$

where s represents the desired level of shape sparsity in the reconstructions. To provide an approximate solution to the combinatorial minimization problem (11), in the sequel we present a relaxation strategy to maintain sufficient smoothness of the cost for

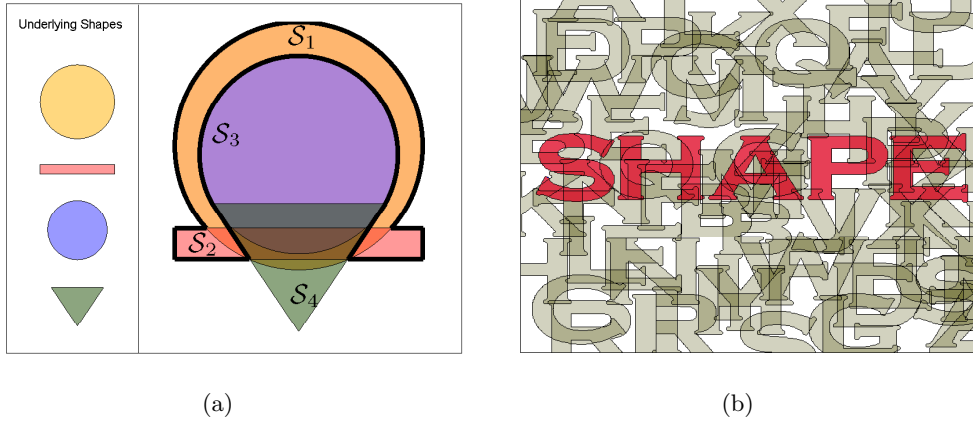


Figure 1: (a) Reconstruction of an “Omega”-shaped region by applying basic set operations on simpler geometries (two circles, a rectangle and a triangle): in this case the desired shape can be written as $\Omega = (\mathcal{S}_1 \cup \mathcal{S}_2) \setminus (\mathcal{S}_3 \cup \mathcal{S}_4)$ (b) Choosing the right components of the desired shape among a dictionary of candidates: in this case extracting the word “SHAPE” from a collection of randomly placed characters

gradient type optimization techniques to become applicable. We then provide a sparse reconstruction algorithm associated with nonlinear costs to employ a sparse number of dictionary elements in the final shape representation.

2.1 Pseudo-Logical Shape Interaction

In a recent work, Aghasi *et al.* provided a rather general view of parametric level set methods for inverse problems. As an example of a parametric form, they proposed a basis expansion of smooth compactly supported radial basis functions (named as bumps) with adaptive dilation and center points. They brought into attention a so called “pseudo-logical” property of this class of functions that we generalize its notion to broader applicability and something not specific to smooth radial bumps.

Consider a given shape $\mathcal{S} \subset D \subset \mathbb{R}^q$. We use the terminology “knoll” for a Lipschitz continuous function $\psi_{\mathcal{S}} : D \rightarrow [0, \infty)$ such that

$$\begin{cases} \psi_{\mathcal{S}}(x) > 0 & x \in \text{int}(\mathcal{S}) \\ \psi_{\mathcal{S}}(x) = 0 & x \in D \setminus \mathcal{S} \end{cases}, \quad (12)$$

with $\text{int}(\mathcal{S})$ denoting the interior of \mathcal{S} . Intuitively, a knoll $\psi_{\mathcal{S}}$ takes positive values inside \mathcal{S} and vanishes outside \mathcal{S} (and on the boundaries as a result of continuity). It can be easily inferred that for two knolls $\psi_{\mathcal{S}_1}$ and $\psi_{\mathcal{S}_2}$ corresponding to shapes \mathcal{S}_1 and \mathcal{S}_2 , the

following facts are true:

$$\text{supp}(\psi_{\mathcal{S}_1} + \psi_{\mathcal{S}_2}) = \mathcal{S}_1 \cup \mathcal{S}_2 \quad (13)$$

and

$$\lim_{\alpha \rightarrow +\infty} \mathcal{D}(\text{supp}^+(\psi_{\mathcal{S}_1} - \alpha\psi_{\mathcal{S}_2}), \mathcal{S}_1 \setminus \mathcal{S}_2) = 0. \quad (14)$$

Here $\text{supp}^+(\cdot)$ denotes the positive support, where the function takes values greater than zero and $\mathcal{D}(\cdot, \cdot)$ is a measure of dissimilarity between two shapes which basically vanishes for identical shapes (see [19] for examples of this measure). The basic message is summing up two knolls would imply a union operation on their supports, and subtracting a knoll of large weight $\alpha \gg 1$ from another knoll approximates applying relative complement on their positive supports.

The idea may be easily incorporated with the notation of level sets by momentarily employing a $c > 0$ level set instead of the zero level set. This unusual lifting is due to the compact support of the knolls, as using a zero level set causes ambiguity in identifying the underlying shape boundaries. In this context, for small values $c > 0$, the c -level set of $\psi_{\mathcal{S}_1} + \psi_{\mathcal{S}_2}$ approximately represents $\mathcal{S}_1 \cup \mathcal{S}_2$ and for large weights $\alpha \gg 1$, the c -level set of $\psi_{\mathcal{S}_1} - \alpha\psi_{\mathcal{S}_2}$ approximately represents $\mathcal{S}_1 \setminus \mathcal{S}_2$ (see Fig 2). Reverting to (10), a more general corollary states that for a parametric shape defined as

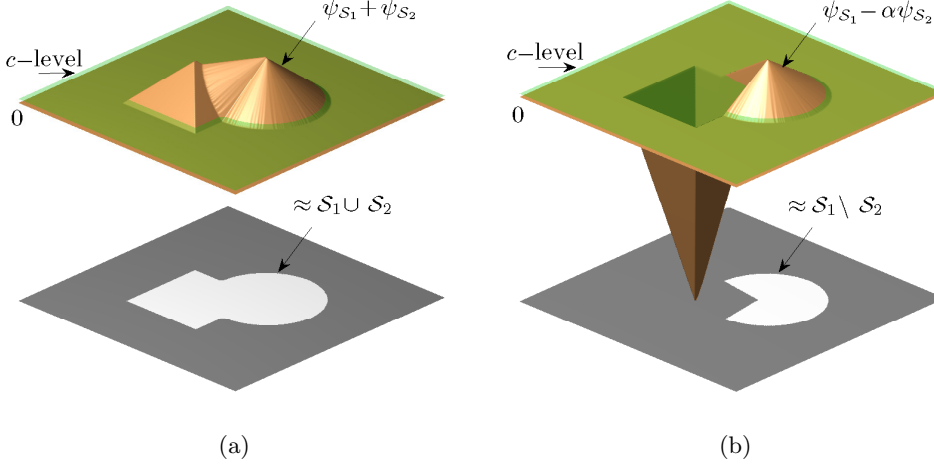


Figure 2: The pseudo-logical behavior of two knolls $\psi_{\mathcal{S}_1}$ and $\psi_{\mathcal{S}_2}$ by considering a close to zero level set c (a) The c -level set of $\psi_{\mathcal{S}_1} + \psi_{\mathcal{S}_2}$ approximately represents $\mathcal{S}_1 \cup \mathcal{S}_2$ (b) When $\alpha \gg 1$ the c -level set of $\psi_{\mathcal{S}_1} - \alpha\psi_{\mathcal{S}_2}$ approximately represents $\mathcal{S}_1 \setminus \mathcal{S}_2$

$$\Omega_{\mathcal{I}_{\oplus}, \mathcal{I}_{\ominus}}^{\alpha} \triangleq \text{supp}^+ \left(\sum_{i \in \mathcal{I}_{\oplus}} \alpha_i \psi_{\mathcal{S}_i}(x) - \sum_{j \in \mathcal{I}_{\ominus}} \alpha_j \psi_{\mathcal{S}_j}(x) - c \right), \quad (15)$$

making

$$c/\alpha_i \rightarrow 0^+ \text{ and } \alpha_j/\alpha_i \rightarrow +\infty, \quad (\text{for all } i \in \mathcal{I}_\oplus, j \in \mathcal{I}_\ominus), \quad (16)$$

results

$$\mathcal{D}(\Omega_{\mathcal{I}_\oplus, \mathcal{I}_\ominus}^\alpha, \Omega_{\mathcal{I}_\oplus, \mathcal{I}_\ominus}) \rightarrow 0. \quad (17)$$

In other words, by considering the c -level set, the parametric function $\sum_{i=1}^{n_d} \alpha_i \psi_{\mathcal{S}_i}(x)$ is capable of producing shapes arbitrarily close to $\Omega_{\mathcal{I}_\oplus, \mathcal{I}_\ominus}$. To conform with prevalent level set methods we henceforth consider the zero level set by transferring the lifting parameter c into the level set formulation as

$$\phi(x, \alpha) = -c + \sum_{i=1}^{n_d} \alpha_i \psi_{\mathcal{S}_i}(x). \quad (18)$$

Of course combining the shapes by varying the coefficients α in (18) brings continuum into the problem. In other words, instead of solving a combinatorial problem and searching among a certain number of possibilities, our search is performed in a subspace spanned by $\psi_{\mathcal{S}_i}(x)$, which contains elements that can represent shapes arbitrarily close to $\Omega_{\mathcal{I}_\oplus, \mathcal{I}_\ominus}$. Searching in this subspace not only relaxes the combinatorial problem and makes it capable of applying descent search methods, also provides the option of exploring shapes that are not among $\Omega_{\mathcal{I}_\oplus, \mathcal{I}_\ominus}$ possibilities (Fig 3).

2.2 Generating Knolls from Given Shapes

Earlier in Section 1.1, we pointed out the use of signed distance functions (SDF) in re-initialization of traditional level sets. In this context at every time step of the Hamilton-Jacobi equation the level set function is re-initialized as an SDF of the form

$$\rho_{\mathcal{S}}(x) = \begin{cases} d(x, \mathcal{C}) & x \in \mathcal{S} \\ -d(x, \mathcal{C}) & x \in D \setminus \mathcal{S} \end{cases}, \quad (19)$$

where \mathcal{S} is the underlying shape and $d(x, \mathcal{C})$ is the distance between the point and the shape boundaries \mathcal{C} . A similar idea may be used to uniquely generate knolls from a given shape. More specifically a knoll may be formed as

$$\psi_{\mathcal{S}}(x) = \begin{cases} d(x, \mathcal{C}) & x \in \mathcal{S} \\ 0 & x \in D \setminus \mathcal{S} \end{cases}, \quad (20)$$

which basically implies $\psi_{\mathcal{S}}(x) = \rho_{\mathcal{S}}(x)^+$. Besides their widespread use in the level set community, our intention of using distance functions is to provide a rather basic expression for the knolls and benefit the fact that shapes at the c -level sets of such knolls uniformly inherit many geometrical features from the original shape (see Fig 4). Moreover, since the c -level set of a knoll $\psi_{\mathcal{S}}$ (or equivalently the zero level set of $\psi_{\mathcal{S}} - c$)

is slightly smaller than \mathcal{S} , in a level set representation as (18), knolls may be generated with slightly larger supports, e.g., $\psi_{\mathcal{S}}(x) = (\rho_{\mathcal{S}}(x) + c)^+$. By this choice, reconstructing exact elements of the dictionary does not require having very large weights associated with their knolls.

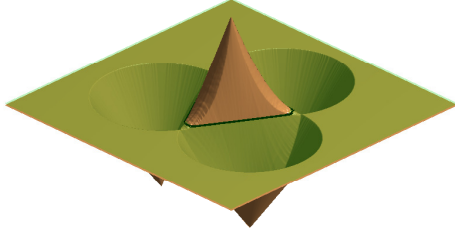


Figure 3: Interaction of four identical circular knolls with different signs which gives rise to a triangular region with almost straight sides. The resulting shape is certainly not in $\Omega_{\mathcal{L}_{\Phi}, \mathcal{L}_{\Theta}}$ possibilities

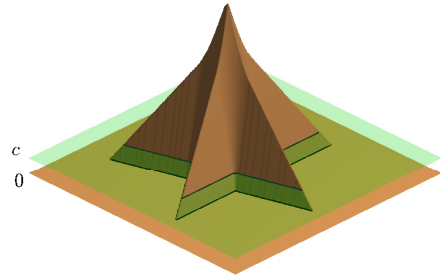


Figure 4: A knoll corresponding to a star-shaped region. The shape resulted at some c -level set near zero inherits many geometrical features of the knoll's support such as the general structure and corners

3 Sparse Reconstruction

We so far discussed how considering a parametric level set as (18) relaxes (11) into a minimization of the form

$$\tilde{\alpha} = \arg \min_{\|\alpha\|_0 \leq s} \mathcal{E}(\alpha), \quad (21)$$

where $\mathcal{E}(\alpha) = \mathcal{E}(\Omega_{\phi(x, \alpha)})$. We are certainly interested in sparse solutions of α to avoid shape redundancy and ill-conditioning associated with large dictionaries. In general \mathcal{E} is a nonlinear and non-convex function of α , we however assume it to be a sufficiently smooth function of α .

In the sequel we provide a general overview of sparse recovery techniques in finding solutions of underdetermined linear systems. We then use this notion to develop our sparsity promoting algorithm applicable to a large class of functionals appeared in imaging applications.

3.1 Background on Sparse Recovery Techniques

Finding sparse solutions of linear systems is a broad area of research in imaging science [28, 33, 43]. For an underdetermined linear system, $\mathbf{A}\boldsymbol{\alpha} = \mathbf{b}$, the main underlying problem is

$$\text{minimize } \|\boldsymbol{\alpha}\|_0 \quad \text{s.t.} \quad \mathbf{A}\boldsymbol{\alpha} = \mathbf{b}. \quad (22)$$

The matrix \mathbf{A} is m -by- n where $m \ll n$ and \mathbf{b} is a vector of length m . Problem (22) is in general a hard combinatorial problem. However, it was brought into attention that relaxing (22) by replacing $\|\boldsymbol{\alpha}\|_0$ with $\|\boldsymbol{\alpha}\|_1$, known as the *basis pursuit* (BP) problem [17], can still result in sparse solutions. Moreover, under certain conditions on \mathbf{A} , a BP solution perfectly coincides with the solution of (22) [12, 13, 22]. In case of noisy observations, the linear equality in (22) is replaced with a least squares inequality as

$$\text{minimize } \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\boldsymbol{\alpha} - \mathbf{b}\|_2 \leq \sigma, \quad (23)$$

where σ is the noise level. This convex minimization is known as the basis pursuit denoising (BPDN) problem. We consider the BP problem as a specific case of BPDN when $\sigma = 0$.

Another convex problem earlier used to induce sparsity on the solutions of a linear system is the *least absolute shrinkage and selection operator* (Lasso) [45], phrased as

$$\text{minimize } \|\mathbf{A}\boldsymbol{\alpha} - \mathbf{b}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_1 \leq \tau. \quad (24)$$

For a certain value $\tau = \tau_\sigma$, problem (24) becomes equivalent to (23), however, in general τ_σ cannot be determined a priori. To promote sparsity, in most applications solving (23) is more desirable than solving (24). This is mainly due to the fact that a good estimate of the noise in the measurements is more likely to be available than prior knowledge of τ , the ℓ_1 -magnitude of a sparse solution.

A variety of solution strategies may be taken for both the BPDN and Lasso problems (e.g., see [11, 26, 34, 35, 45]). Relatively speaking, the smooth cost associated with the Lasso makes it an easier problem to approach compared to the BPDN which involves a non-smooth cost minimization. Minimization of smooth costs over convex sets may be efficiently handled using methods such as spectral projected gradient (SPG) [7].

We briefly overview the SPG-based technique proposed by van den Berg and Friedlander in [47], who show that a numerically efficient way of solving the BPDN problem is by solving a series of Lasso problems. This technique plays a major role in developing the algorithm that we put forth later in this paper. Their method also allows use of \mathbf{A} as a linear operator with no explicit matrix form which suits our general representation.

Following [47], for a given $\tau \geq 0$, a single parameter function $\varphi(\tau)$ is defined as

$$\varphi(\tau) \triangleq \|\mathbf{b} - \mathbf{A}\boldsymbol{\alpha}_\tau\|_2, \quad (25)$$

where $\boldsymbol{\alpha}_\tau$ is a solution of (24) for the given τ . It should be noted that multiple solutions to the Lasso may exist, however, problem's convexity requires all of them to generate an equal error term $\varphi(\tau)$. It is shown in [47] that $\varphi(\cdot)$ is a nonincreasing convex function which is continuously differentiable and

$$\varphi'(\tau) = \frac{d\varphi}{d\tau} = -\frac{\|\mathbf{A}^T(\mathbf{b} - \mathbf{A}\boldsymbol{\alpha}_\tau)\|_\infty}{\|\mathbf{b} - \mathbf{A}\boldsymbol{\alpha}_\tau\|_2}. \quad (26)$$

The graph of $\varphi(\tau)$ in terms of τ provides an optimal trade-off between the residual error $\|\mathbf{b} - \mathbf{A}\boldsymbol{\alpha}_\tau\|_2$ and the ℓ_1 -norm of the solution, forming the Pareto curve (Fig 5). The Lasso and BPDN problems are basically two different characterizations of this curve. In other words, for a given σ a solution to

$$\varphi(\tau) = \sigma, \quad (27)$$

returns the value of τ_σ that makes solutions of BPDN and Lasso problems coincide. As a matter of fact solving (27) does not require full access to the Pareto curve. The convex nature of the problem and a closed form expression for $\varphi'(\tau)$ are sufficient to employ root finding techniques and acquire τ_σ . In this context, a Newton iterative process $\tau_{\ell+1} = \tau_\ell + \Delta\tau_\ell$, where

$$\Delta\tau_\ell = \frac{\sigma - \varphi(\tau_\ell)}{\varphi'(\tau_\ell)}, \quad \ell = 0, 1, \dots \quad (28)$$

is capable of generating a sequence of Lasso parameters τ_ℓ that superlinearly converge to τ_σ . This result makes Lasso the central tool in solving the BPDN problem. More specifically, for a given $\sigma \geq 0$, a solution to the BPDN problem is acquired by solving a series of Lasso problems parameterized by τ_ℓ and ultimately arriving at a Lasso that is parameterized by τ_σ , which basically solves the BPDN problem. To solve the underlying Lasso problems efficiently authors in [47] employ the SPG technique detailed in [7].

To address a broader class of problems, specifically in a variational framework, in our future formulations we look beyond matrix representatives of \mathbf{A} . We consider a more general case that $A : \mathbb{R}^n \rightarrow \mathbb{S}$ is a bounded linear operator and \mathbb{S} represents a real Hilbert space². An example of \mathbb{S} , other than \mathbb{R}^m , is the space of bounded continuous functions $C(\mathbb{R})$. This generalization only requires slight modifications to the above overview of sparse recovery techniques, basically by replacing $\|\cdot\|_2$ with $\|\cdot\|_{\mathbb{S}}$, the norm induced by the Hilbert space \mathbb{S} , and using the adjoint operator A^* in place of \mathbf{A}^T . The sparsity constraints applied to $\boldsymbol{\alpha}$ remain intact as the domain of A is still \mathbb{R}^n . For linear systems, the technique proposed in [47] makes such generalization possible and conveniently handles cases that A is not explicitly available as a matrix.

²We therefore keep the notation general and do not use a bold syntax (representative of a matrix form) for the linear operator and elements of \mathbb{S} .

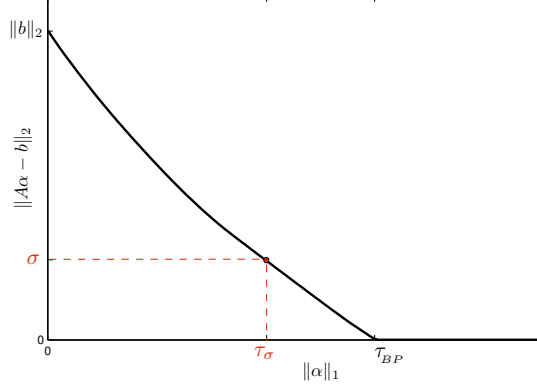


Figure 5: A typical Pareto curve corresponding to the Lasso and BPDN problems. For a given σ , the corresponding value τ_σ makes Lasso and BPDN problems share solution. For values of τ larger than τ_{BP} (the ℓ_1 -norm of the BP solution), the residual $\|A\alpha - b\|_2$ is zero.

3.2 A More General Sparsity Constrained Optimization

The majority of sparsity-constrained optimization techniques are merely applicable to linear systems and corresponding quadratic costs. We are however interested in sparse stationary points of $\mathcal{E}(\alpha)$ in (21), which is not generally a quadratic function of α . This is a more complex problem and a new area of research which still requires further development. The current available techniques are mainly in the form of (21) for which we are required to know the degree of sparsity, s , prior to the minimization [3, 4, 8].

For the purpose of this paper, we consider a particular form of $\mathcal{E}(\alpha)$ applicable to a large class of imaging problems. We then provide a sparsity promoting minimization scheme that takes a rather short iterative process to converge and does not require prior assignment of s . Accordingly, for a given real Hilbert space \mathbb{S} , we consider $\mathcal{E}(\alpha)$ to take a form (or a finite sum of forms) as

$$\mathcal{E}(\alpha) = \|\mathcal{G}(\alpha)\|_{\mathbb{S}}^2 = \langle \mathcal{G}(\alpha), \mathcal{G}(\alpha) \rangle_{\mathbb{S}}, \quad (29)$$

where $\mathcal{G} : \mathbb{R}^n \rightarrow \mathbb{S}$ is a Fréchet differentiable map applied to α . The classic compressed sensing problem can be interpreted as a special case of (29) when $\mathcal{G}(\cdot)$ is an affine function of α , i.e., $\mathcal{G}(\alpha) = A\alpha - b$. In the following we present two classes of imaging applications that follow the form (29) and will be later considered in Section 4.

3.2.1 Shape-Based Inverse Problems

For this class of problems, in the basic case, the spatial parameter of interest u is modelled as

$$u(x) = \begin{cases} u_{in} & x \in \Omega \\ u_{ex} & x \in D \setminus \Omega \end{cases}, \quad (30)$$

where u_{in} corresponds to the texture value of an inclusion (the shape) and u_{ex} represents the background texture. The texture values may be represented by scalars or low order parametric models which may/may not be known a priori [1]. Using a parametric function $\phi(x, \alpha)$ to characterize the shape by its zero level set we may rewrite the parameter of interest as

$$u(x, \alpha) = u_{in}H(\phi(x, \alpha)) + u_{ex}(1 - H(\phi(x, \alpha))), \quad (31)$$

where $H(\cdot)$ is the heaviside function, usually replaced with a smooth approximation $H_{rg}(\cdot)$ to maintain differentiability [1]. The inverse problem may now be formulated as minimizing the model-data mismatch to acquire the shape parameters α , i.e.,

$$\min_{\alpha} \|v - \mathcal{M}(u(x, \alpha))\|_{\mathbb{S}_v}^2. \quad (32)$$

For a smooth physical model \mathcal{M} that relates the parameter of interest u to the observations, the underlying inversion cost may be interpreted as a functional of the form (29) when $\mathcal{G}(\alpha) = v - \mathcal{M}(u(x, \alpha))$.

3.2.2 Image Segmentation

A well known variation form to segment an image $u(x)$ into two disjoint regions Ω and $D \setminus \Omega$ corresponds to a functional

$$\mathcal{E}(\Omega) = \mathcal{E}_r(\Omega) + \int_{\Omega} r_{in}(u(x)) \, dx + \int_{D \setminus \Omega} r_{ex}(u(x)) \, dx, \quad (33)$$

where $\mathcal{E}_r(\Omega)$ represents some regularity constraints on the segmented regions (e.g., smoothness and compactness [14, 16]) and $r_{in}(\cdot) \geq 0$ and $r_{ex}(\cdot) \geq 0$ are some inhomogeneity measures of u in each region. A well known instance of such measures is observed in the work by Chan and Vese [16], which suggests $r_{in}(u) = (u(x) - \tilde{u}_{in})^2$ and $r_{ex}(u) = (u(x) - \tilde{u}_{ex})^2$. Here \tilde{u}_{in} and \tilde{u}_{ex} are scalars representing the average pixel value within each region. Another example is using a maximum likelihood approach for the pixel intensities to arrive at

$$r_{\chi}(u) = -\log(p_{\chi}(u(x))), \quad \chi : in, ex, \quad (34)$$

where $p_{in}(u)$ and $p_{ex}(u)$ are pixel intensity distributions [20, 41]. The scalar parameters \tilde{u}_{in} , \tilde{u}_{ex} and the distributions $p_{in}(\cdot)$, $p_{ex}(\cdot)$ may be known a priori or may be estimated in

the course of segmentation. Considering the image dependent terms in (33), a parametric level set function may be employed to form the segmentation functional

$$\begin{aligned}\mathcal{E}(\boldsymbol{\alpha}) &= \int_D r_{in}(u(x))H(\phi(x, \boldsymbol{\alpha})) \, dx \\ &\quad + \int_D r_{ex}(u(x))(1 - H(\phi(x, \boldsymbol{\alpha}))) \, dx.\end{aligned}\tag{35}$$

We note that in using a parametric level set technique the regularity constraints may either be neglected thanks to the intrinsic smoothness of the parametric function [27], or in our case taken into consideration via the knolls sparsity. The energy functional in (35) may be written as sum of two functionals $\mathcal{E}_{in}(\boldsymbol{\alpha})$ and $\mathcal{E}_{ex}(\boldsymbol{\alpha})$ in form of (29) using $\mathcal{G}_{in}(\boldsymbol{\alpha}) = \sqrt{r_{in}(u)H(\phi(x, \boldsymbol{\alpha}))}$ and $\mathcal{G}_{ex}(\boldsymbol{\alpha}) = \sqrt{r_{ex}(u)(1 - H(\phi(x, \boldsymbol{\alpha})))}$. Employing a smooth function $H_{rg}(\cdot) : \mathbb{R} \rightarrow [0, 1]$ to approximate the Heaviside function can guarantee smoothness of $\mathcal{G}_{in}(\cdot)$ and $\mathcal{G}_{ex}(\cdot)$ in $\boldsymbol{\alpha}$.

3.3 A Sparse Nonlinear Minimization Technique

To minimize a sufficiently smooth cost function $\mathcal{E}(\boldsymbol{\alpha}) : \mathbb{R}^n \rightarrow \mathbb{R}$, a well known iterative scheme is Newton's method. Starting from an initial vector $\boldsymbol{\alpha}_0$, Newton's method proceeds by generating $\boldsymbol{\alpha}_k$ vectors ($k = 1, 2, \dots$) that progressively reduce the cost to reach a minima. Having $\boldsymbol{\alpha}_k$ available, the successive vector is written as $\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \boldsymbol{\delta}_k$ (or a multiple of $\boldsymbol{\delta}_k$), where the step is determined by minimizing the second order Taylor approximation of the cost around $\boldsymbol{\alpha}_k$:

$$\boldsymbol{\delta}_k = \arg \min_{\boldsymbol{\delta}} \mathcal{E}(\boldsymbol{\alpha}_k) + \mathcal{J}_{\mathcal{E}}(\boldsymbol{\alpha}_k)\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^T \mathcal{H}_{\mathcal{E}}(\boldsymbol{\alpha}_k)\boldsymbol{\delta}.\tag{36}$$

Here $\mathcal{J}_{\mathcal{E}}(\boldsymbol{\alpha}_k)$ denotes the Jacobian vector at $\boldsymbol{\alpha}_k$ and $\mathcal{H}_{\mathcal{E}}(\boldsymbol{\alpha}_k)$ is the corresponding Hessian matrix.

A positive definite Hessian guarantees existence of a minima for (36) which leads to the closed form expression $\boldsymbol{\delta}_k = (\mathcal{H}_{\mathcal{E}}(\boldsymbol{\alpha}_k))^{-1}\mathcal{J}_{\mathcal{E}}(\boldsymbol{\alpha}_k)$. Most Newton type techniques either impose such positivity constraint on the Hessian or approximate it with an at least positive semi-definite matrix to guarantee convexity of the underlying cost [21].

To consider applying a Newton type technique to the energy functional in (29), using calculus of operators [1, 29], we have

$$\mathcal{J}_{\mathcal{E}}(\boldsymbol{\alpha})\boldsymbol{\delta} = 2 \langle \mathcal{G}'(\boldsymbol{\alpha})\boldsymbol{\delta}, \mathcal{G}(\boldsymbol{\alpha}) \rangle_{\mathbb{S}},\tag{37}$$

and

$$\frac{1}{2}\boldsymbol{\delta}^T \mathcal{H}_{\mathcal{E}}(\boldsymbol{\alpha})\boldsymbol{\delta} = \langle \mathcal{G}'(\boldsymbol{\alpha})\boldsymbol{\delta}, \mathcal{G}'(\boldsymbol{\alpha})\boldsymbol{\delta} \rangle_{\mathbb{S}} + \langle \mathcal{G}''(\boldsymbol{\alpha})[\boldsymbol{\delta}, \boldsymbol{\delta}], \mathcal{G}(\boldsymbol{\alpha}) \rangle_{\mathbb{S}}.\tag{38}$$

Here $\mathcal{G}'(\boldsymbol{\alpha})[\cdot] : \mathbb{R}^n \rightarrow \mathbb{S}$ is a linear operator representing the first order Fréchet derivative of \mathcal{G} at $\boldsymbol{\alpha}$ and $\mathcal{G}''(\boldsymbol{\alpha})[\cdot, \cdot] : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{S}$ is a bilinear operator that represents the second

order Fréchet derivative of \mathcal{G} . When \mathbb{S} is taken to be \mathbb{R}^m , \mathcal{G}' and \mathcal{G}'' are simply the Jacobian matrix and the Hessian tensor of \mathcal{G} . For a positive semi-definite approximation to $\mathcal{H}_{\mathcal{E}}$, in Gauss-Newton methods [21] the inner product term containing \mathcal{G}'' is neglected in (38). Therefore the underlying quadratic cost corresponding to (36) becomes

$$\begin{aligned} \|\mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2 + 2 \langle \mathcal{G}'(\alpha_k)\delta, \mathcal{G}(\alpha_k) \rangle_{\mathbb{S}} + \|\mathcal{G}'(\alpha_k)\delta\|_{\mathbb{S}}^2 \\ = \|\mathcal{G}'(\alpha_k)\delta + \mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2, \end{aligned}$$

forming a convex function of δ . Although the underlying quadratic cost is convex, the linear operator \mathcal{G}' may be ill conditioned and some type of regularization may be required to obtain δ_k . For instance Levenberg-Marquardt algorithm is a variant of Gauss-Newton technique that uses a Tikhonov type regularizer to minimize the underlying quadratic cost [21].

We however consider a different regularizer which is capable of promoting sparsity on consecutive estimates of α_k . More specifically having α_k available we regularize the subproblem by promoting sparsity on the next potential estimate α_{k+1} through a BPDN problem, i.e.,

$$\delta_k = \arg \min_{\delta} \quad \|\alpha_k + \delta\|_1 \quad \text{s.t.} \quad \|\mathcal{G}'(\alpha_k)\delta + \mathcal{G}(\alpha_k)\|_{\mathbb{S}} \leq \sigma. \quad (39)$$

This is certainly a translated version of the BPDN problem and a simple change of variable as $\theta = \alpha_k + \delta$ converts it to the standard form (23). As δ is expected to be very small about an accumulation point, a reasonable choice of σ is an estimate of $\|\mathcal{G}(\alpha)\|_{\mathbb{S}}^2$ near a minima. In Appendix A we show that when $\sigma = \min_{\alpha} \|\mathcal{G}(\alpha)\|_{\mathbb{S}}^2$, or even more locally when $\sigma \leq \|\mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2$, the direction δ_k acquired in (39) is a descent direction.

In the case of inverse problems, σ may be an estimate of noise power in the measurements. Nevertheless, if such estimate is not available, a substitute for (39) would be a parameter free BP problem

$$\begin{aligned} \delta_k = \arg \min_{\delta} \quad & \|\alpha_k + \delta\|_1 \\ \text{s.t.} \quad & \mathcal{G}'^*(\alpha_k)\mathcal{G}'(\alpha_k)\delta + \mathcal{G}'^*(\alpha_k)\mathcal{G}(\alpha_k) = 0, \end{aligned} \quad (40)$$

corresponding to the classic Gauss-Newton equation with an additional sparsity constraint. If $\mathcal{G}'(\alpha_k)\delta + \mathcal{G}(\alpha_k) = 0$ has a class of solutions, which is very likely for an ill-conditioned linear system, this equality constraint may be replaced with the one in (40) and make it equivalent to the case $\sigma = 0$ in (39).

As stated in Section 3.1, at every iteration a minimizer to (39) may be determined via a sequence of Lasso solves

$$\begin{aligned} \delta_{k,\ell} = \arg \min_{\delta} \quad & \|\mathcal{G}'(\alpha_k)\delta + \mathcal{G}(\alpha_k)\|_{\mathbb{S}} \\ \text{s.t.} \quad & \|\alpha_k + \delta\|_1 \leq \tau_{k,\ell}, \end{aligned} \quad (41)$$

where for $\ell = 0, 1, 2, \dots$ the ℓ_1 radius $\tau_{k,\ell}$ is updated according to (28) and sequentially $\lim_{\ell \rightarrow \infty} \delta_{k,\ell} = \delta_k$. Once the k -th iteration is complete, a similar Lasso search is performed along the Pareto curve corresponding to $\|\mathcal{G}'(\alpha_{k+1})\delta + \mathcal{G}(\alpha_{k+1})\|_{\mathbb{S}}$ to find δ_{k+1} .

In determining the consecutive directions δ_k , it is computationally desirable to perform incomplete iterations on ℓ and make a more accurate search for δ_k when α_k approaches a sparse solution. For this purpose, it is also shown in Appendix A that if for some ℓ , $\tau_{k,\ell} \geq \|\alpha_k\|_1$, the corresponding step, $\delta_{k,\ell}$, is a descent direction and can be used as a step to reduce the nonlinear cost.

Also, for small values of k that α_k may be far from a sparse solution, $\tau_{k,\ell}$ may tend to very large values because of the high cardinality of α_k . To prevent this, the values $\tau_{k,\ell}$ may be controlled by a loose threshold value τ_{mx} and the iterations on ℓ may be broken if this threshold is reached. The descent property of the incomplete step for this case is also inferred by a similar argument as the one just mentioned. These two strategies provide us with an indication of when we are eligible to break the ℓ -iterations and still be certain that the descent property of resulting direction is maintained.

Algorithm 1 provides a detailed picture of the proposed minimization scheme where a line search technique is employed to warrant a proper convergence behavior. We specifically use Armijo rule to determine the step size, which requires the reduction in the cost to be sufficiently large [6]. Parameters γ and β correspond to this reduction and usually take values as $\gamma \in [10^{-5}, 0.1]$ and $\beta \in [0.1, 0.5]$, detailed in [6]. The values ε_1 and ε_2 control the size of updates in each block and are very small numbers in the order of machine precision. The value of τ_{mx} may be taken as a loose overestimate of the ℓ_1 -norm of the solution. Finally α_0 is the initialization of unknown parameters and τ_0 is an initial ℓ_1 -ball radius which may arbitrarily taken to be zero or very small.

3.4 Employing an Asymmetric Norm

In the sparsity promoting algorithm proposed we used the ℓ_1 norm as a convex relaxation to the ℓ_0 norm. This is a common technique in compressed sensing promoting sparsity on the reconstructions thanks to the sharp vertices of the ℓ_1 ball. However, in general an ℓ_0 minimization only targets the cardinality of a vector while an ℓ_1 minimization also takes into account the vector component values.

For a parametric representation as (18), Fig 6(a) shows reconstruction of an L-shaped region where a basic ℓ_1 minimization does not necessarily provide the best sparse solution. Considering a dictionary that contains all shapes $\mathcal{S}_1, \dots, \mathcal{S}_6$ shown in the figure, in one case the L-shape is reconstructed via the union of four rectangular knolls while a similar reconstruction is possible by applying the relative complement among only two knolls. When $\|\alpha\|_1$, the ℓ_1 norm of knoll coefficients, is used to indicate the level of sparsity, the latter case may have a larger ℓ_1 norm although only two shapes are exploited. This is because of the relatively large negative coefficients that are required to simulate a

Algorithm 1 A sparsity promoting Gauss-Newton algorithm

```
input  $\sigma$ ;  
set  $\gamma, \beta, \varepsilon_1, \varepsilon_2$  and  $\tau_{mx}$ ;  
 $\alpha := \alpha_0$ ;  
 $\tau := \tau_0$ ;  
 $found := \mathbf{false}$ ;  
while  $\sim found$  do  
   $stepfound := \mathbf{false}$ ;  
  while  $\sim stepfound$  do  
     $\delta := \arg \min_{\eta} \|\mathcal{G}'(\alpha)\eta + \mathcal{G}(\alpha)\|_{\mathbb{S}} \text{ s.t. } \|\alpha + \eta\|_1 \leq \tau$ ;  
     $\phi'_\tau := -\|\mathcal{G}'^*(\alpha)[\mathcal{G}'(\alpha)\delta + \mathcal{G}(\alpha)]\|_{\infty} / \|\mathcal{G}'(\alpha)\delta + \mathcal{G}(\alpha)\|_{\mathbb{S}}$ ;  
     $\Delta\tau := (\sigma - \|\mathcal{G}'(\alpha)\delta + \mathcal{G}(\alpha)\|_{\mathbb{S}}) / \phi'_\tau$ ;  
     $\tau := \tau + \Delta\tau$ ; (alternatively,  $\tau := \min(\tau + \Delta\tau, \tau_{mx})$ );  
     $stepfound := (\tau \geq \|\alpha\|_1) \vee (\Delta\tau \leq \varepsilon_1)$ ;  
  end while  
  while  $\mathcal{E}(\alpha) - \mathcal{E}(\alpha + \delta) < -\gamma \mathcal{J}_{\mathcal{E}}(\alpha)\delta$  do  
     $\delta := \beta * \delta$ ;  
  end while  
   $\alpha := \alpha + \delta$ ;  
   $found := (\|\delta\|_2 \leq \varepsilon_2)$ ;  
end while
```

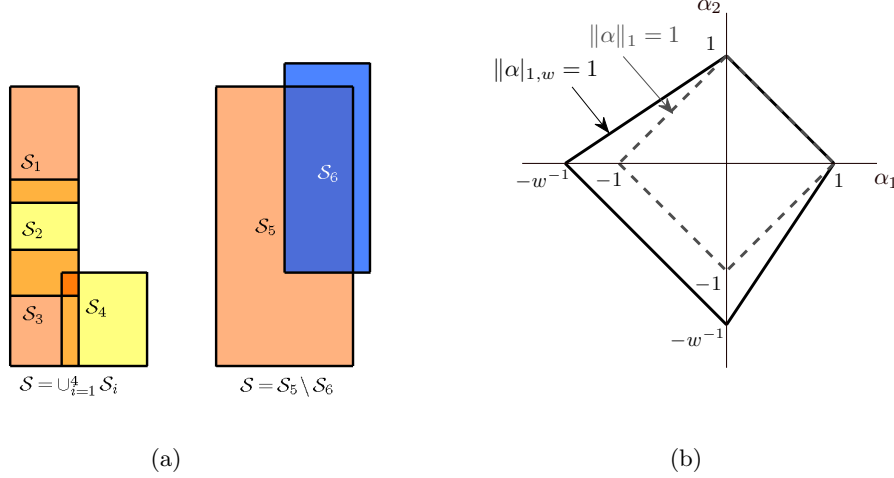


Figure 6: Employing an asymmetric ℓ_1 norm: (a) Reconstructing an L-shaped region by union of four rectangles (left) and applying the relative complement among two rectangles (right); (b) The ℓ_1 ball and the asymmetric convex ball corresponding to $\|\alpha\|_{1,w}$

relative complement operation (revisit Fig 2).

One way to remedy such phenomenon is to consider an asymmetric ℓ_1 norm defined as

$$\|\alpha\|_{1,w} \triangleq \sum_{\{i: \alpha_i \geq 0\}} |\alpha_i| + \sum_{\{j: \alpha_j < 0\}} w|\alpha_j|, \quad (42)$$

where w is a constant scalar. When $w \in (0, 1)$, this representation allows the appearance of larger negative coefficients by weighting them less in the absolute sum. Additionally, setting $w \gg 1$ promotes positivity on the coefficients which is desirable in some applications that we will consider in next section.

Clearly $\|\cdot\|_{1,w}$ is not formally a norm as it violates the positive homogeneity of a norm, however, the corresponding ball is still convex (see Fig 6(b)). A generalization of [47] reported in [48] allows replacing the ℓ_1 constraint with any convex constraints. As stated in [48] for this generalization the infinity norm in (26) needs to be replaced with its *polar* (reducing to the dual norm when the constraint is a norm), which in this case is the asymmetric infinity norm

$$\|\alpha\|_{\infty, w^{-1}} \triangleq \max_{\substack{i: \alpha_i \geq 0 \\ j: \alpha_j < 0}} \{|\alpha_i|, w^{-1}|\alpha_j|\}. \quad (43)$$

To more conveniently facilitate our method with this feature, we would note that in a

lower implementation level, a main component of the SPG method in [47] is performing iterative projections of the form

$$\tilde{\alpha}^\perp = \arg \min_{\alpha} \|\alpha - \tilde{\alpha}\|_{\mathbb{S}} \quad s.t. \quad \|\alpha\|_{1,w} \leq \tau. \quad (44)$$

The asymmetric norm maybe written as $\|\alpha\|_{1,w} = \|\mathbf{D}(\alpha)\alpha\|_1$ where $\mathbf{D}(\alpha)$ is a diagonal matrix with entries

$$\mathbf{D}(\alpha)_{i,i} = \begin{cases} 1 & \alpha_i \geq 0 \\ -w & \alpha_i < 0 \end{cases}. \quad (45)$$

As we have shown in Appendix B, solution of (44) coincides with the solution of the weighted ℓ_1 minimization problem

$$\min_{\alpha} \|\alpha - \tilde{\alpha}\|_{\mathbb{S}} \quad s.t. \quad \|\mathbf{D}(\tilde{\alpha})\alpha\|_1 \leq \tau. \quad (46)$$

This fact is well demonstrated in Fig 7. Based on this argument we may replace the projection onto an asymmetric ball problem with a weighted ℓ_1 minimization, which is straightforward.

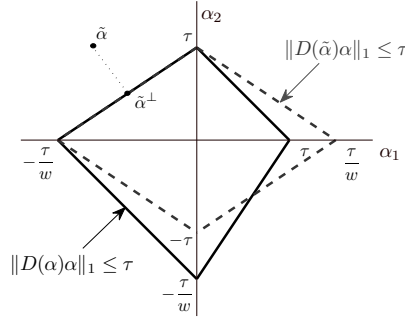


Figure 7: Coincidence of the projections onto the asymmetric balls $\|\alpha\|_{1,w} \leq \tau$ and $\|\mathbf{D}(\tilde{\alpha})\alpha\|_1 \leq \tau$. The two balls share side in the second quadrant where $\tilde{\alpha}$ is located.

4 Examples

In this section we examine the proposed technique in various imaging applications. For the examples presented, we use the approximate Heaviside function suggested in [1] as

$$H_\epsilon(x) = \begin{cases} 1 & x > \epsilon \\ 0 & x < -\epsilon \\ \frac{1}{2} + \frac{x}{2\epsilon} + \frac{1}{2\pi} \sin\left(\frac{\pi x}{\epsilon}\right) & |x| \leq \epsilon, \end{cases} \quad (47)$$

which gives rise to a compactly supported approximation of $\delta_\epsilon(x) = H'_\epsilon(x)$. As discussed in [1] in the context of inverse problems, this choice results in a form of narrow-banding in the evolution of the parameters, that will also be discussed briefly in segmentation examples. The parameter ϵ determines the width of the narrow-band region, whereas smaller values of ϵ brings sharper transitions into the reconstructions at the expense of slower evolution [1]. The choice of the lifting parameter c is quite arbitrary as the proposed shape-based problem is relative, i.e., simultaneously scaling c and α_i coefficients in (18) does not change the zero level set shape. In other words, taking larger values of c would tend to larger values of α_i in the reconstructions. Since in general the shapes in the dictionary may significantly vary in size, to have knoll basis terms of comparable magnitudes, we normalize each knoll to its maximum value. For the majority of examples we simply take $c = 0.1$ and $\epsilon = 0.05$ and initialize the algorithm from a rather random state. Based on the scaling property between c and final α_i values, τ_{mx} may be taken as a large multiple of c , e.g., $\tau_{mx} = 50c$.

4.1 Image Segmentation

As discussed in Section 3.2.2, for a binary Chan-Vese segmentation the energy functional finds the form of (35) which maybe written as the sum of internal and external energies:

$$\mathcal{E}(\boldsymbol{\alpha}) = \|\mathcal{G}_{in}(\boldsymbol{\alpha})\|_{\mathbb{S}}^2 + \|\mathcal{G}_{ex}(\boldsymbol{\alpha})\|_{\mathbb{S}}^2. \quad (48)$$

Accordingly a reasonable representation for $\mathcal{G}(\boldsymbol{\alpha})$ would be

$$\mathcal{G}(\boldsymbol{\alpha}) = \mathcal{G}_{in}(\boldsymbol{\alpha}) + \mathbf{i} \mathcal{G}_{ex}(\boldsymbol{\alpha}), \quad (49)$$

where $\mathbf{i} = \sqrt{-1}$. The underlying Hilbert space would be $\mathbb{S} = L^2(D)$ with the corresponding inner product

$$\langle s_1(x), s_2(x) \rangle_{L^2(D)} = \int_D s_1(x) \overline{s_2(x)} dx, \quad (50)$$

where $\overline{s_2(x)}$ represents the complex conjugate of $s_2(x)$. Clearly, applying our proposed algorithm to the segmentation problem requires having $\mathcal{G}'(\boldsymbol{\alpha})[\cdot]$ and its adjoint form. For a given vector $\boldsymbol{\eta} \in \mathbb{R}^{n_d}$, a formal derivation of (49) with respect to $\boldsymbol{\alpha}$ and applying the resulting Jacobian operator to $\boldsymbol{\eta}$ yields

$$\mathcal{G}'(\boldsymbol{\alpha})\boldsymbol{\eta} = f(x, \boldsymbol{\alpha}) \sum_i^{n_d} \eta_i \psi_{\mathcal{S}_i}(x), \quad (51)$$

where

$$f(x, \boldsymbol{\alpha}) = \begin{cases} \frac{1}{2} \left(\sqrt{\frac{r_{in}(u)}{H_\epsilon(\phi)}} - \mathbf{i} \sqrt{\frac{r_{ex}(u)}{1-H_\epsilon(\phi)}} \right) \delta_\epsilon(\phi) & |\phi| < \epsilon \\ 0 & else. \end{cases} \quad (52)$$

From a numerical perspective, the linear operator $\mathcal{G}'(\boldsymbol{\alpha})$ is a matrix with columns $f(x, \boldsymbol{\alpha})\psi_{\mathcal{S}_i}(x)$. Based on this argument, the adjoint operator may be easily deduced. For a function $\psi \in L^2(D)$ the linear operation $\mathcal{G}'^*(\boldsymbol{\alpha})\psi$ produces an array of size n_d where

$$[\mathcal{G}'^*(\boldsymbol{\alpha})\psi]_i = \left\langle \overline{f(x, \boldsymbol{\alpha})}\psi_{\mathcal{S}_i}(x), \psi(x) \right\rangle_{L^2(D)}. \quad (53)$$

To maintain generality, we preferred providing the implicit forms (51) and (53) for $\mathcal{G}'(\boldsymbol{\alpha})$ and $\mathcal{G}'^*(\boldsymbol{\alpha})$, however, for pixelated images of moderate size a matrix may be assigned to each linear operator.

Equation (52) shows that $f(x, \boldsymbol{\alpha})$ is a compactly supported function. Similarly, $\psi_{\mathcal{S}_i}(x)$ is compactly supported, and therefore when $\text{supp}(\delta_\epsilon(\phi)) \cap \text{supp}(\psi_{\mathcal{S}_i}) = \emptyset$, the corresponding columns of $\mathcal{G}'(\boldsymbol{\alpha})$ would become zero. In determining a descent direction, such columns may be neglected as they would not affect the result. In other words, in iteratively updating the knoll coefficients α_i , the only knolls that are updated are the ones that intersect the narrow-band $\delta(\phi)$ at that iteration. This numerical advantage is basically the *narrow-banding* feature associated with the proposed parametric representation and extensively described in [1] in the context of inverse problems.

The technical details presented may be employed to address variants of Chan-Vese segmentation in different applications. Here we consider two applications that could be addressed through a sparse shape recovery.

4.1.1 Image Segmentation with Missing Pixels

A challenging problem associated with image segmentation arises when some of the pixels are missing or part of the image is occluded. This limits the analysis to only a subset of image pixels, while the segmentation needs to be performed globally. In this example we show that prior information about the geometry of objects in the image can lead to perform a completion to missing pixels when a sparse shape recovery is considered.

Fig 8 shows a noisy reference image that is occluded with rectangular patches. Additionally, random pixels of the image are discarded resulting in overall pixel loss of 50% and 80% shown in Figures 8(b) and 8(c). These two figures will be the subject of our segmentation.

Our prior information about the geometry of objects in the image is reflected in the choice of dictionary elements. To build up the dictionary we make use of four basic shapes: a circle, square, triangle and an ellipse, shown in Fig 8(d). We would note that the triangle in the reference image is the upper diagonal portion of a square while the one used in the dictionary is a lower portion.

A slight modification of the formulation presented above needs to be considered for this limited observation problem. More specifically, considering $D' \subset D$ to be the available portion of the imaging domain D , the Hilbert space considered would be $L^2(D')$ where

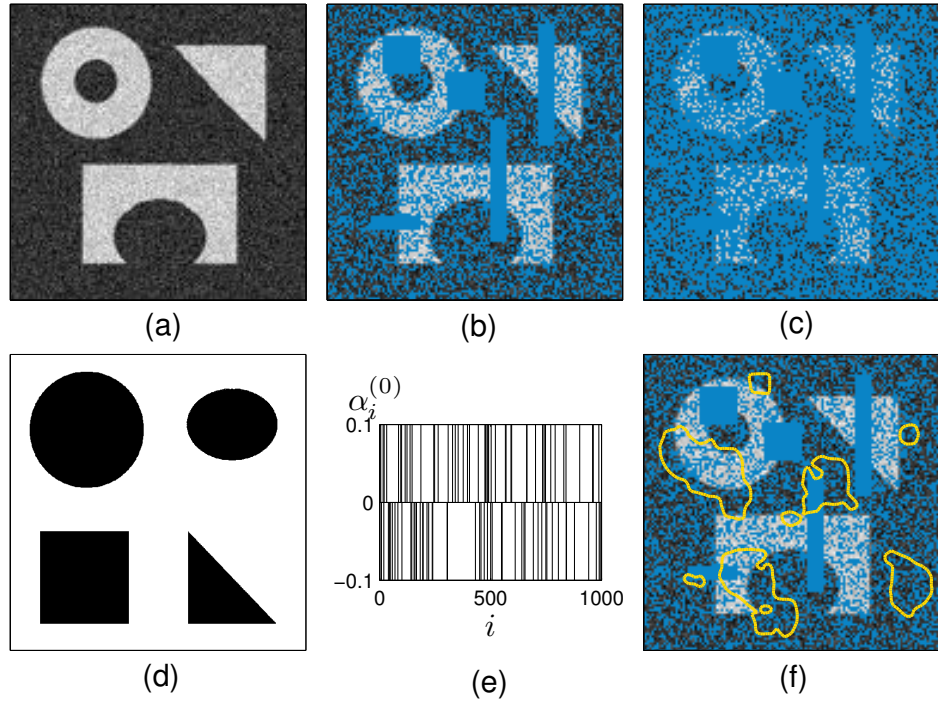


Figure 8: (a) Reference image (b) A test image with more than 50% of the pixels missing (blue regions show the missing regions) (c) A test image with more than 80% of the pixels missing (d) Shapes used to build up the dictionary: 1000 instances of these four shapes with different sizes are placed throughout the imaging domain (e) Initial values of α_i for $i = 1, \dots, 1000$ (f) The resulting initial contour for the given initial α . The segmentation results are shown in Fig 9.

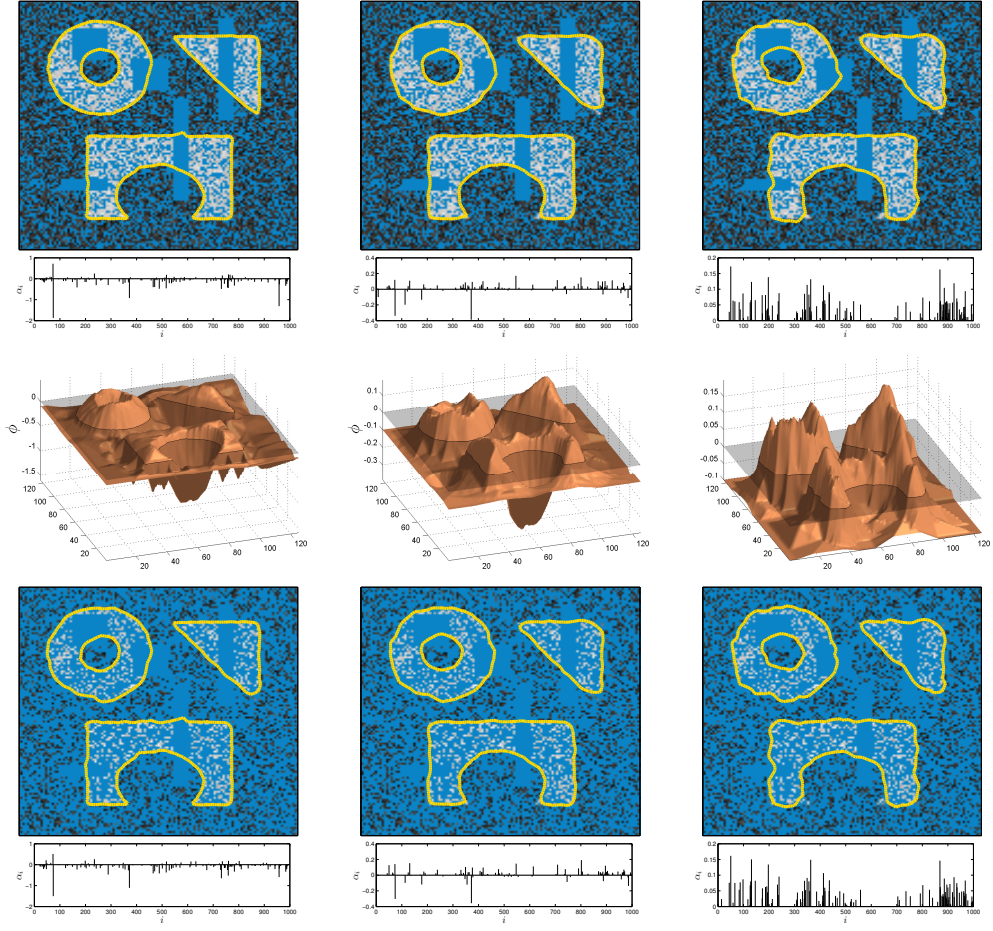


Figure 9: First row shows the results of segmentation on 50% missing-pixel image, from left to right for $w = 0.1$, $w = 1$ and $w = 10$. Below every result the final sparse vector α is shown. Second row are the level set functions corresponding to each segmentation above it. Third row shows an identical scenario as the first row applied to the image with 80% missing-pixel

the observation takes place. The knoll functions, however, are constructed and defined globally in D .

By considering different size and placements of the four basic shapes, a total of $n_d = 1000$ shapes are assigned to the dictionary. To initialize the algorithm we randomly apply weights of +1 and -1 to 100 knolls as shown in Fig 8(e). This assignment results the initial shape shown in Fig 8(f). A basic Chan-Vese segmentation cost is considered for which the values \tilde{u}_{in} and \tilde{u}_{ext} are occasionally updated following the formulation in [16].

The first column of Fig 9 shows the successful segmentation results on the two test images that are obtained for $w = 0.1$. Although significant parts of the images are missing, in both cases the proposed algorithm has made a reasonable segmentation job. To highlight the advantage of considering an asymmetric norm, results are also shown for the basic ℓ_1 case of $w = 1$ and an inappropriate choice of $w = 10$ which basically abandons set minus operation. It can be seen that although considering smaller values for w increases the size of feasible region and requires solving the problem in a larger domain, it pays off by promoting the set minus operation and using a sparser set of shapes in a more efficient manner. This contrast is clearly observable by comparing the results of first and third columns. Basically, in the third column knolls are pushed to take positive weights which degrades the reconstruction by exclusively promoting the union operation. Thanks to the appearance of the set minus operation, the segmentation results in the first column have sharper corners and smoother sides closer to the truth while using relatively less number of shapes in the dictionary. It is worth noting that all segmentations converged the steady state in less than 15 iterations.

4.1.2 Text Recognition: Breaking a Basic Captcha

Captcha (Completely Automated Public Turing test to tell Computers and Humans Apart) is a well-known test in the world of computers that ensures that response to a query is generated by a human [50]. For this test, the client is asked to read and type the word in an image, where the characters are placed in an unusual manner not recognizable for the machine. Since the underlying components of the image are still characters, this maybe considered as a challenging sparse shape reconstruction problem. The underlying shapes in the dictionary may constitute a dense set of possibilities for the character shapes that may appear in a Captcha image. Of course a denser dictionary increases the chances of identifying the characters correctly.

As a proof of concept, in this example we consider an attack on the rather basic Captcha image of size 63×160 pixels shown in Fig 10(a) which represents the word “*ShaPE*”. The letters in the image may take different case, size, rotations and overlap as the case for our Captcha image. To build up the character dictionary we assume knowing the font type a priori and consider all 52 uppercase and lowercase letters of the English

alphabet. The displacement possibilities considered for each character are 64 points on a regular grid in the image. At every point, five different rotations of the character and two different font sizes are considered. This setup leads to $n_d = 52 \times 64 \times 5 \times 2 = 33280$ shapes in the dictionary. Of course for a more realistic scenario, font possibilities, deformations and more number of size, displacement and rotational variations may be considered for every letter. We however keep the problem small to be tractable with a desktop computer. Since the main purpose of this problem is identifying shapes in the dictionary that appear somewhere in the image, w may be chosen to be large (10 for our simulations) to push the algorithm on only considering the union operation.

As the first experiment we consider the case that exact character shapes in the Captcha are present in the dictionary. As before, the algorithm is initialized with positive and negative weights for a random subset of α_i coefficients which corresponds to the initial shape contour in Fig 10(b). The classic Chan-Vese cost functional is again considered with the texture parameters updated occasionally. The segmentation result after 12 iterations is shown in Fig 10(c) and in Fig 10(d) we have shown the coefficients values. A simple index map relating the weight indices to the alphabetical representation of the shape indicates that the reconstructed image is composed of five main characters, precisely matching with the word inside the image.

For the second example we consider the Captcha in Fig 11(a) where the letters in the image do not precisely match the dictionary elements. The closest elements of the dictionary are depicted with colored contours on the same image. This problem may be still solved using a multi-stage refinement strategy. More specifically, a first segmentation attempt shown in Fig 11(b) with the underlying weights shown in Fig 11(c) does not provide a sparse solution with prominent weights as before. Instead, in this case there are several candidates, the top 20 of which are shown in Fig 11(c). Based on the first stage results, a new dictionary maybe built up, with much a much smaller number of elements, only containing denser size, rotation, displacement variants of the shapes listed in the top 20 list. Performing the segmentation process again, this time provides us with the weights shown in Fig 11(e) where the striking weights only correspond to the letters in the Captcha. The “echo” effect corresponding to appearance of multiple prominent weights of the same letter is due to the close distance between the shapes in the new dictionary. Of course when the indexing mechanism takes into account the character positions, all such echoes correspond to close positions and maybe unified as a single final conclusion about the letter in that region of the image.

4.2 Medical Imaging: X-ray Computed Tomography (CT)

As a well known linear inverse problem, in this section we examine the method in reconstructing CT images. For a mono-energetic CT, X-ray photons are transmitted through the test medium and measured at the opposite side. If the medium has an attenuation

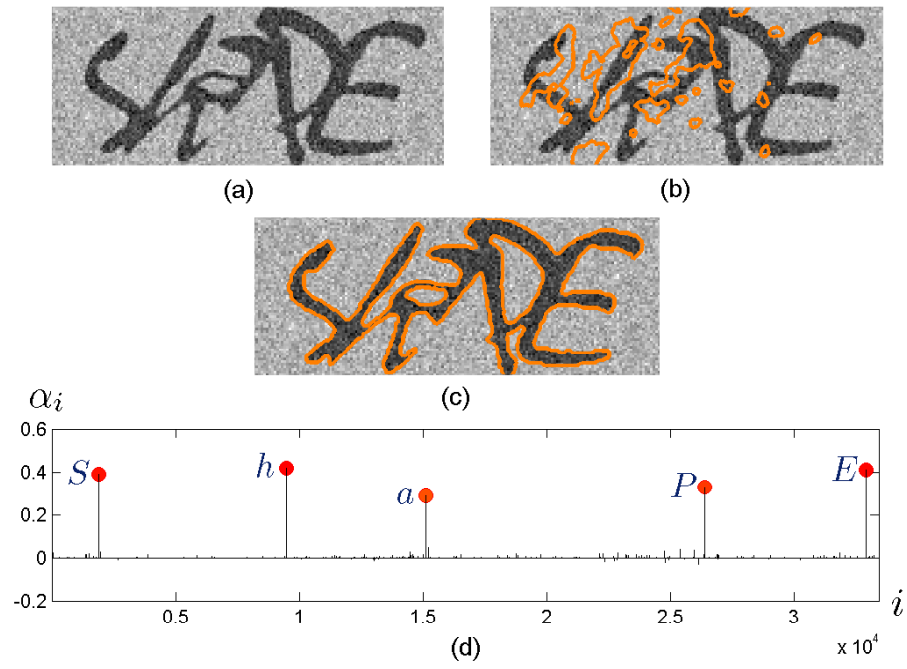


Figure 10: (a) A Captcha image (b) The initial contour corresponding to the random initialization of the α_i coefficients (c) The segmentation result shown with colored contour (d) The values of reconstructed weights and an indication of the letter each index corresponds to

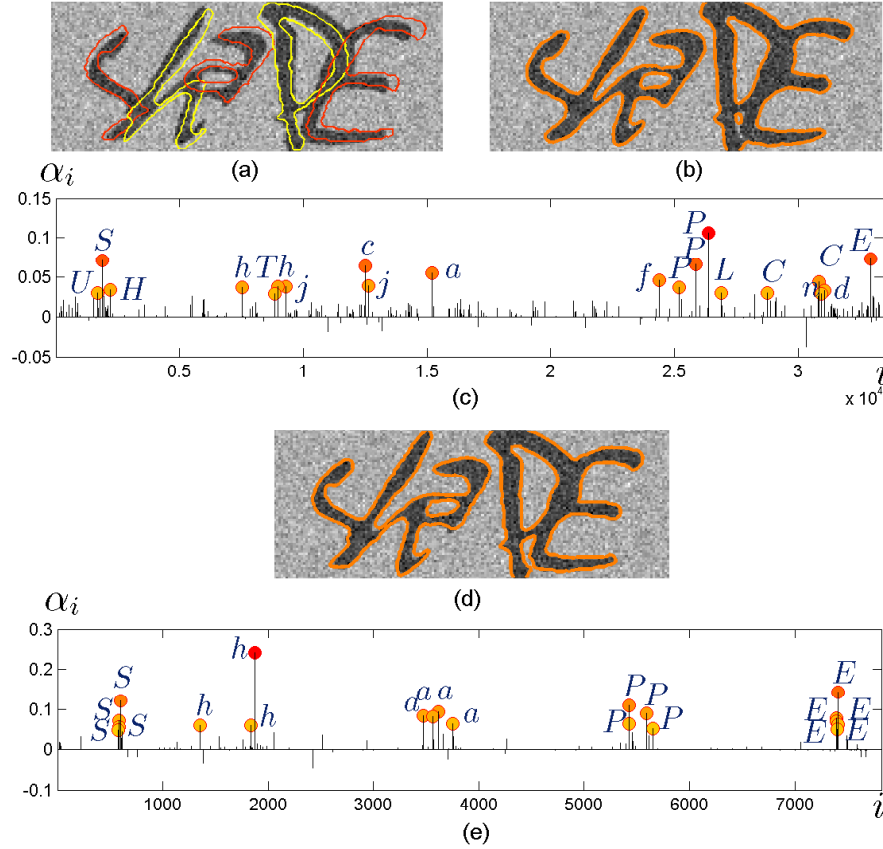


Figure 11: (a) A Captcha image where the letters in the image do not match with the elements of the dictionary; the closest elements of the dictionary are shown with colored contours (b) A first stage reconstruction results (c) Reconstructed weights corresponding to the first stage segmentation; the top 20 candidates are indicated with the corresponding letter (d) Segmentation results after refining the dictionary according to the first stage results (e) The weights corresponding to the second stage reconstruction indicating the letter each index corresponds to

profile $\mu(x)$ the number of photons measured would ideally be

$$\lambda_m = \lambda_T \exp\left(-\int_{\mathcal{L}_m} \mu(x) dx\right), \quad m = 1, \dots, M, \quad (54)$$

where λ_m is the photon count measured at m -th receiver, λ_T is the blank scan photon count and \mathcal{L}_m is the line through which the ray travels. An easier way of interpreting the measurements is reading the values

$$v_m = -\log \frac{\lambda_m}{\lambda_T} = \int_{\mathcal{L}_m} \mu(x) dx, \quad (55)$$

at each sensor which basically relates the attenuation map to the measurements via a Radon transform. In practice however, the measurements are corrupted with Poisson noise, i.e.,

$$\tilde{\lambda}_m = \text{Pois}(\lambda_m) \quad (56)$$

and the true measurements are $\tilde{v}_m = -\log(\tilde{\lambda}_m/\lambda_T)$. A second order approximation to the log likelihood function indicates that [44]

$$\log p(\tilde{\mathbf{v}}|\mu) \approx -\frac{1}{2}(\tilde{\mathbf{v}} - \mathcal{R}\mu)^T \mathbf{D}(\tilde{\mathbf{v}} - \mathcal{R}\mu) + h(\tilde{\mathbf{v}}) \quad (57)$$

where $\mathbf{D} = \text{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_M)$, \mathcal{R} is the linear Radon type transform that maps μ to the ideal measurements and $h(\cdot)$ is a function dependent only on the data vector $\tilde{\mathbf{v}}$. To apply the proposed algorithm to this modality the attenuation profile is parameterized as $\mu(x, \boldsymbol{\alpha})$, and the residual operator is written as $\mathcal{G}(\boldsymbol{\alpha}) = \mathcal{R}\mu(x, \boldsymbol{\alpha}) - \tilde{\mathbf{v}}$. For a maximum likelihood estimate of the parameters, the underlying cost takes the form $\|\mathcal{G}(\boldsymbol{\alpha})\|_{\mathbf{D}}^2$ where the inner product in the measurement space is defined as $\langle \mathbf{s}_1, \mathbf{s}_2 \rangle_{\mathbf{D}} = \mathbf{s}_1^T \mathbf{D} \mathbf{s}_2$ for $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^M$. Based on the linearity of the CT model the Jacobian operator applied to a vector $\boldsymbol{\eta}$ may be written as

$$\mathcal{G}'(\boldsymbol{\alpha})\boldsymbol{\eta} = \sum_i \eta_i \frac{\partial \mathcal{G}(\boldsymbol{\alpha})}{\partial \alpha_i} = \sum_i \eta_i \mathcal{R} \frac{\partial \mu}{\partial \alpha_i} = \mathcal{R} \sum_i \eta_i \frac{\partial \mu}{\partial \alpha_i} \quad (58)$$

and the adjoint operator applied to a vector $\boldsymbol{\psi} \in \mathbb{R}^M$ takes the form of

$$\mathcal{G}'^*(\boldsymbol{\alpha})\boldsymbol{\psi} = \langle \mathcal{R} \frac{\partial \mu}{\partial \alpha_i}, \boldsymbol{\psi} \rangle_{\mathbf{D}}. \quad (59)$$

Equations (58) and (59) are in general the key components of applying Algorithm 1 to this problem.

Fig 12(a) shows the true attenuation profile of a 200×200 pixel chest test image. To generate the X-ray data the rays are emitted at 60 equispaced angles between 0 and 180

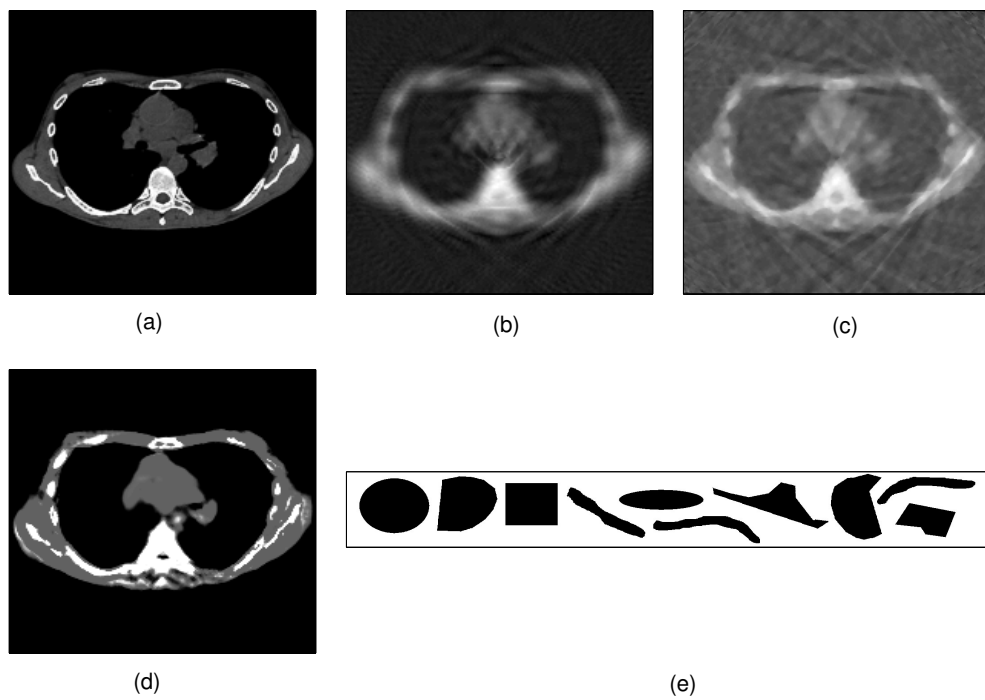


Figure 12: (a) The true attenuation profile of a chest image (b) Inversion using FBP (c) TV inversion (d) Sparse shape inversion results (e) The basic shapes used in the dictionaries

degrees. At each angle an average of 38 photon rays are transmitted through the image. A total of 2280 measurements are obtained among which 1499 are the ones crossing the chest and containing useful information. The blank scan photon count is $\lambda_T = 4 \times 10^6$ and the average photon energies are 50 KeV. Beside considering the Poisson noise, 1% zero mean Gaussian noise is also added to the data to model the inaccuracy in the measurement readings. This problem setup poses a challenging ill-posed problem which is rather hard to approach. In Fig 12(b) we have shown the filtered back projection (FBP) results for this data set. Fig 12(c) also shows the reconstruction results using a total variation (TV) approach employing the ℓ_1 -magic package [11].

For the purpose of shape representation we consider a multi-phase level set approach [49]. More specifically to invert for both the soft tissue and bone geometries along with the blank space we consider using two level sets $\phi_1(x, \boldsymbol{\alpha}^{(1)})$ and $\phi_2(x, \boldsymbol{\alpha}^{(2)})$ as

$$\begin{aligned} \mu(x, \boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}) &= \mu_a + (\mu_s - \mu_a)H_\epsilon(\phi_1(\boldsymbol{\alpha}^{(1)})) \\ &\quad + (\mu_b - \mu_s)H_\epsilon(\phi_1(\boldsymbol{\alpha}^{(1)}))H_\epsilon(\phi_2(\boldsymbol{\alpha}^{(2)})) \end{aligned}$$

where μ_a , μ_s and μ_b are respectively the average attenuation values for air, soft tissue and bone. Considering the average density of each material, at 50 KeV photon energy, the average attenuation values for the air, soft tissue and bone are approximately 2.7×10^{-4} , 0.2 and 0.7 cm^{-1} [30]. We use two shape dictionaries $\mathfrak{D}^{(1)}$ and $\mathfrak{D}^{(2)}$. The former contains $n_d^{(1)} = 5346$ shapes and the latter is in hold of $n_d^{(2)} = 6156$ shapes. The basic elements used in the dictionaries are those shown in Fig 12(e). Elements of $\mathfrak{D}^{(1)}$ are chosen to have larger sized shapes as ϕ_1 is mainly in charge of representing the soft tissue. It is however worth noting that both dictionaries share identical shapes. The shapes are placed all over the imaging domain specially in places that there are chances of objects being present. Plugging the sensitivity relations

$$\frac{\partial \mu}{\partial \alpha_i^{(1)}} = \psi_{\mathcal{S}_i}^{(1)} \left((\mu_s - \mu_a)\delta_\epsilon(\phi_1) + (\mu_b - \mu_s)\delta_\epsilon(\phi_1)H_\epsilon(\phi_2) \right),$$

and

$$\frac{\partial \mu}{\partial \alpha_i^{(2)}} = \psi_{\mathcal{S}_i}^{(2)} \left((\mu_b - \mu_s)H_\epsilon(\phi_1)\delta_\epsilon(\phi_2) \right),$$

into (58) and (59) provides the necessary components in using Algorithm 1 to invert for the level set coefficients. The process of determining a descent direction may be performed in a coordinate descent fashion by alternatively updating the weights for the first and second level set at each iteration. Fig 12(d) shows the result of our inversion. Through a sparse composition of shapes we have been able to reconstruct a reasonable estimate of true image. Considering the image to be a composition of shapes provides a better pose to the problem compared to the case of considering it as an image with piecewise constant regions suitable for a TV inversion.

4.3 Resistance Tomography: An Archaeological Problem

As the last example we consider an archaeological application of imaging subsurface tombs using electrical resistance tomography (ERT). The severely ill-posed nature of the problem as well as the limitations in placing the measuring sensors make this problem very challenging specially when the structures are closely spaced [25].

In this technique electric current is injected into the ground and some sensors measure the resulting potential on different regions of the imaging domain. Based on these potential measurements an inverse problem is solved to reconstruct the profile of electrical conductivity. The governing physics may be described as

$$\begin{aligned}\nabla \cdot (\sigma \nabla \rho) &= j \quad \text{in } D, \\ \xi_1 \sigma \frac{\partial \rho}{\partial \mathbf{n}} + \xi_2 \rho &= 0 \quad \text{on } \partial D,\end{aligned}\tag{60}$$

where j denotes the pattern of injected current, σ is the conductivity and ρ is the resulting potential. Functions ξ_1 and ξ_2 are functions defined on the boundary of the imaging domain and are in charge of imposing appropriate boundary conditions [2].

To maintain simplicity, consider ρ_s to be the potential resulted from a point source current $j(x) = \delta(x - x_s)$ and $\rho_{s,m}$ representing the measured voltage at points x_m in the domain for $m = 1, \dots, M$. The forward model is a nonlinear operator that maps $\sigma(x)$ to the potential measurements. To apply the proposed technique we need to know about the model Jacobian operator. Using the adjoint field technique, it can be shown that perturbations in the measurements are related to the conductivity perturbations through [2]

$$\delta \rho_{s,m} = \int_D \delta \sigma \nabla \rho_s \cdot \nabla \rho_m dx.\tag{61}$$

Here ρ_m is the potential resulted from placing the point source current at x_m , known as the adjoint field. When a parametric form $\sigma = \sigma(x, \boldsymbol{\alpha})$ is considered for the conductivity, we have $\delta \sigma = \sum_i \delta \alpha_i \partial \sigma / \partial \alpha_i$, where $\delta \alpha_i$ is the perturbation of the i -th element of $\boldsymbol{\alpha}$. Based on this argument

$$\mathcal{G}'_m(\boldsymbol{\alpha})\eta = \int_D \nabla \rho_s \cdot \nabla \rho_m \left(\sum_i \eta_i \frac{\partial \sigma}{\partial \alpha_i} \right) dx.\tag{62}$$

For $\boldsymbol{\alpha} \in \mathbb{R}^{n_d}$, the model Jacobian operator $\mathcal{G}'(\boldsymbol{\alpha})$ is a matrix of size $M \times n_d$ the (m, i) entry of which is $\int_D \partial \sigma / \partial \alpha_i \nabla \rho_s \cdot \nabla \rho_m dx$. Accordingly, the adjoint operator $\mathcal{G}'^*(\boldsymbol{\alpha})$ is simply the transposed matrix in this case. This notion can be extended to more complex cases of running different experiments with different sources, as well as more sophisticated source settings such as electric dipoles.

Fig (13) shows the setup for a field experiment in a region where two closely placed tombs exist. A total of 50 sensors are placed on the ground and 20 experiments are carried out at each experiment using a pair of sensors as the electric dipole (shown with

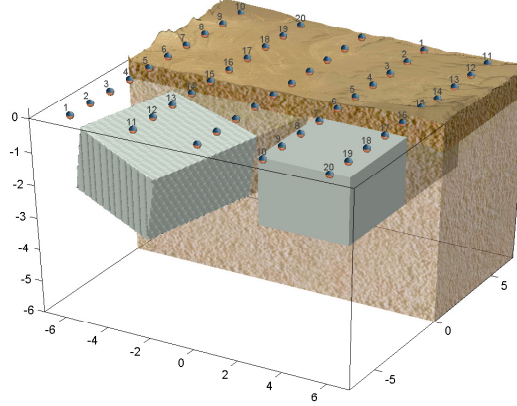


Figure 13: The ERT setting and sensor configuration

identical numbers) and the remaining sensors as the measurement sensors. The data is polluted with 1% white noise. The true cavities are modeled as cubic structures shown in the figure. The structure on the left is slightly tilted in both azimuth and elevation. Following (31) the conductivity distribution is modeled as

$$\sigma(x, \boldsymbol{\alpha}) = \sigma_s + (\sigma_a - \sigma_s)H_\epsilon(\phi(x, \boldsymbol{\alpha})), \quad (63)$$

where σ_a and σ_s are the average known values for the conductivity of air and soil. A total of $n_d = 14157$ shape are used in the dictionary. The dictionary elements are cubic structures distributed all over D , noting that the true cavities are not among the elements of the dictionary. Fig 14(b) shows the initialization of the algorithm and Fig 14(c) is the imaging results after 11 iterations.

Reconstruction results after 43 iterations using conventional level set technique are also shown in Fig 14(e). We would like to note that the conventional level set inversion is initialized with a very good initialization as shown in Fig 14(e). As observable, the proposed technique is able to provide a more accurate profile of the subsurface conductivity and successfully separate the two structures. The counterpart, however, is unable to provide such level of detail and fails to make the separation in deeper regions of the ground that the sensitivity values are lower.

5 Conclusion

The idea presented in this paper may be considered as a new technique in approaching variety of shape-based imaging problems. The main message of this work is changing the

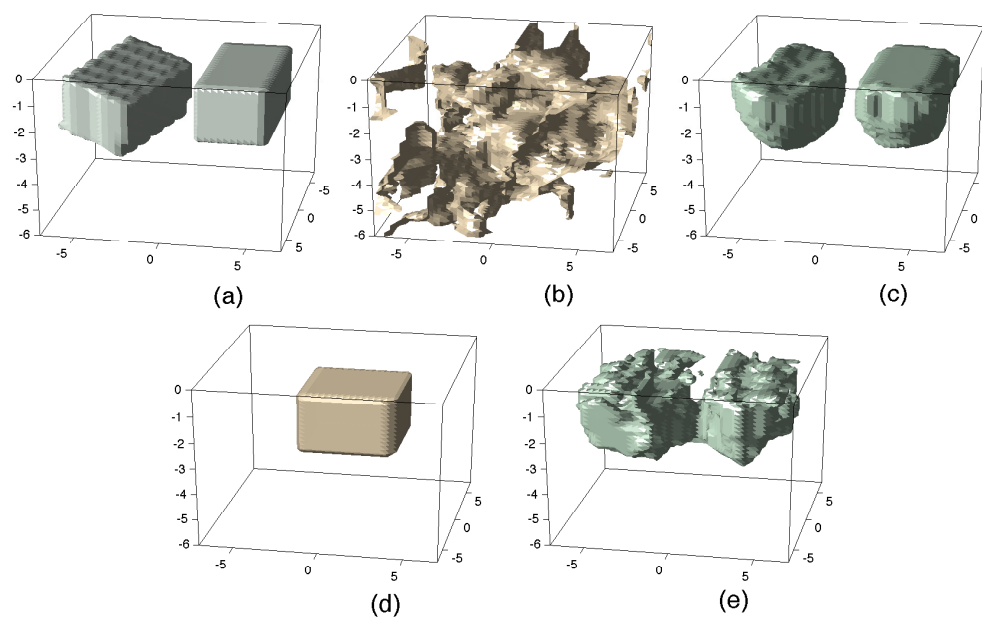


Figure 14: (a) The true tombs (b) Initialization of the algorithm (c) Reconstruction results (d) Initialization of traditional level sets (e) Reconstruction results using traditional level set technique

geometric problem of shape composition into a variational problem that may be more conveniently analyzed. This conversion is performed rather simply through the notion of pseudo-logical property. In fact following this idea, the shape composition problem becomes similar to the classic problem of representing a function with a weighted sum of functional elements from a dictionary. The notion of sparsity comes next as a means of choosing proper elements, however, the shape-based nature of the problem requires struggling with a nonlinear problem. By considering several examples from different applications, we showed that, although a nonlinear problem, the proposed sparsity promoting technique can successfully handle struggling problems with large dictionary elements. However, more general techniques applicable to larger class of functionals without restrictive assumptions such as knowing the sparsity degree a priori are desirable and welcomed. Of course the non-convex nature of the shape-based problem does not allow talking about uniqueness of reconstructions, however, our group is still working on appropriately restricting the problem for which such analysis is possible. This is still an open arena of research that requires exploring new techniques in nonlinear problems with sparsity constraints.

A Descent Property of Proposed Steps

We first show that when σ is sufficiently small a direction acquired from (39) is descent. First a quadratic expansion yields

$$\|\mathcal{G}'(\alpha_k)\delta + \mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2 = \|\mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2 + \mathcal{J}_{\mathcal{E}}(\alpha)\delta + \|\mathcal{G}'(\alpha_k)\delta\|_{\mathbb{S}}^2.$$

As δ_k needs to meet the constraint in (39) we must have $\|\mathcal{G}'(\alpha_k)\delta_k + \mathcal{G}(\alpha_k)\|_{\mathbb{S}} \leq \sigma$, which requires

$$\begin{aligned} \mathcal{J}_{\mathcal{E}}(\alpha_k)\delta_k &\leq \sigma - \|\mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2 - \|\mathcal{G}'(\alpha_k)\delta_k\|_{\mathbb{S}}^2 \\ &\leq \sigma - \|\mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2. \end{aligned} \tag{64}$$

The most right side of (64) is certainly negative for sufficiently small values of σ , when for instance $\sigma = \min_{\alpha} \|\mathcal{G}(\alpha)\|_{\mathbb{S}}^2$ or more locally $\sigma < \|\mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2$. Such choices make δ_k a descent direction for \mathcal{E} at α_k .

We next show that when $\tau_{k,\ell} \geq \|\alpha_k\|_1$, direction $\delta_{k,\ell}$ acquired from (41) is descent. Since $\|\alpha_k + \mathbf{0}\|_1 \leq \tau_{k,\ell}$ and $\delta_{k,\ell}$ is a minima for (41) we need to have

$$\begin{aligned} \|\mathcal{G}'(\alpha_k)\delta_{k,\ell} + \mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2 &\leq \|\mathcal{G}'(\alpha_k)\mathbf{0} + \mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2 \\ &= \|\mathcal{G}(\alpha_k)\|_{\mathbb{S}}^2, \end{aligned} \tag{65}$$

which using (64) simplifies to

$$\mathcal{J}_{\mathcal{E}}(\alpha_k)\delta_{k,\ell} \leq -\|\mathcal{G}'(\alpha_k)\delta_{k,\ell}\|_{\mathbb{S}}^2 \leq 0. \tag{66}$$

B The Equivalent Problem to the Projection onto an Asymmetric ℓ_1 -Ball

Proposition: For $\tilde{\alpha} \in \mathbb{R}^n$ given, consider $\tilde{\alpha}_1^\perp$ to be the projection onto an asymmetric ℓ_1 -ball as

$$\tilde{\alpha}_1^\perp = \arg \min_{\alpha} \|\alpha - \tilde{\alpha}\|_{\mathbb{S}} \quad s.t. \quad \|\alpha\|_{1,w} \leq \tau, \quad (67)$$

and $\tilde{\alpha}_2^\perp$ to be the solution to the weighted ℓ_1 projection problem

$$\tilde{\alpha}_2^\perp = \arg \min_{\alpha} \|\alpha - \tilde{\alpha}\|_{\mathbb{S}} \quad s.t. \quad \|D(\tilde{\alpha})\alpha\|_1 \leq \tau, \quad (68)$$

where the diagonal matrix $D(\tilde{\alpha})$ is defined as (45). Then $\tilde{\alpha}_1^\perp = \tilde{\alpha}_2^\perp$.

Proof: We first note that both problems (67) and (68) are framed as a projection onto a convex set, and in both cases the minima is unique (see Prop B.11 in [6]). Clearly elements of $\tilde{\alpha}_1^\perp$ and $\tilde{\alpha}$ must have identical signs over the support of $\tilde{\alpha}_1^\perp$; otherwise by changing the signs of incompatible elements a feasible point is obtained with a lower cost, which contradicts $\tilde{\alpha}_1^\perp$ being the global minima. A similar argument holds for $\tilde{\alpha}_2^\perp$ and $\tilde{\alpha}$, and so

$$D(\tilde{\alpha}_1^\perp)\tilde{\alpha}_1^\perp = D(\tilde{\alpha})\tilde{\alpha}_1^\perp, \quad (69)$$

and

$$D(\tilde{\alpha}_2^\perp)\tilde{\alpha}_2^\perp = D(\tilde{\alpha})\tilde{\alpha}_2^\perp. \quad (70)$$

Suppose that $\tilde{\alpha}_1^\perp \neq \tilde{\alpha}_2^\perp$. Since $\tilde{\alpha}_2^\perp$ is a solution to (68), it needs to satisfy $\|D(\tilde{\alpha}_2^\perp)\tilde{\alpha}_2^\perp\|_1 \leq \tau$. This result along with (70) and the fact that $\|\alpha\|_{1,w} = \|D(\alpha)\alpha\|_1$, reveal that $\tilde{\alpha}_2^\perp$ is also a feasible point for problem (67). Therefore, based on the strict convexity of the cost

$$\|\tilde{\alpha}_1^\perp - \tilde{\alpha}\|_{\mathbb{S}} < \|\tilde{\alpha}_2^\perp - \tilde{\alpha}\|_{\mathbb{S}}. \quad (71)$$

In a similar fashion equation (69) reveals that $\tilde{\alpha}_1^\perp$ is a feasible point for problem (68) and therefore $\|\tilde{\alpha}_2^\perp - \tilde{\alpha}\|_{\mathbb{S}} < \|\tilde{\alpha}_1^\perp - \tilde{\alpha}\|_{\mathbb{S}}$ which contradicts (71).

References

- [1] A. AGHASI, M. KILMER, AND E.L. MILLER, *Parametric level set methods for inverse problems*, SIAM Journal on Imaging Sciences, 4 (2011), pp. 618–650.
- [2] A. AGHASI AND E.L. MILLER, *Sensitivity calculations for poisson's equation via the adjoint field method*, Geoscience and Remote Sensing Letters, IEEE, 9 (2012), pp. 237–241.

- [3] S. BAHMANI, P. BOUFONOS, AND B. RAJ, *Greedy sparsity-constrained optimization*, in Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on, IEEE, 2011, pp. 1148–1152.
- [4] A. BECK AND Y.C. ELDAR, *Sparsity constrained nonlinear optimization: Optimality conditions and algorithms*, arXiv preprint arXiv:1203.4580, (2012).
- [5] O. BERNARD, D. FRIBOULET, P. THÉVENAZ, AND M. UNSER, *Variational B-spline level-set: a linear filtering approach for fast deformable model evolution*, IEEE Transactions on Image Processing, 18 (2009), pp. 1179–1191.
- [6] D.P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.
- [7] E.G. BIRGIN, J.M. MARTÍNEZ, AND M. RAYDAN, *Nonmonotone spectral projected gradient methods on convex sets*, SIAM Journal on Optimization, 10 (2000), pp. 1196–1211.
- [8] T. BLUMENSATH, *Compressed sensing with nonlinear observations and related nonlinear optimisation problems*, arXiv preprint arXiv:1205.1650, (2012).
- [9] M. BURGER, B. HACKL, AND W. RING, *Incorporating topological derivatives into level set methods*, Journal of Computational Physics, 194 (2004), pp. 344–362.
- [10] M. BURGER AND S.J. OSHER, *A survey on level set methods for inverse problems and optimal design*, European Journal of Applied Mathematics, 16 (2005), pp. 263–301.
- [11] E. CANDÈS AND J. ROMBERG, *l1-magic: Recovery of sparse signals via convex programming*, URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf, 4 (2005).
- [12] E.J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, Information Theory, IEEE Transactions on, 52 (2006), pp. 489–509.
- [13] E.J. CANDÈS, J.K. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on pure and applied mathematics, 59 (2006), pp. 1207–1223.
- [14] V. CASELLES, R. KIMMEL, AND G. SAPIRO, *Geodesic active contours*, International journal of computer vision, 22 (1997), pp. 61–79.
- [15] T.F. CHAN, J. SHEN, AND L. VESE, *Variational pde models in image processing*, Notices of AMS, 50 (2003).

- [16] T.F. CHAN AND L.A. VESE, *Active contours without edges*, Image Processing, IEEE Transactions on, 10 (2001), pp. 266–277.
- [17] S.S. CHEN, D.L. DONOHO, AND M.A. SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM review, (2001), pp. 129–159.
- [18] E.T. CHUNG, T.F. CHAN, AND X.C. TAI, *Electrical impedance tomography using level set representation and total variational regularization*, Journal of Computational Physics, 205 (2005), pp. 357–372.
- [19] D. CREMERS, S.J. OSHER, AND S. SOATTO, *Kernel density estimation and intrinsic alignment for shape priors in level set segmentation*, International Journal of Computer Vision, 69 (2006), pp. 335–351.
- [20] D. CREMERS, M. ROUSSON, AND R. DERICHE, *A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape*, International journal of computer vision, 72 (2007), pp. 195–215.
- [21] J.E. DENNIS AND R.B. SCHNABEL, *Numerical methods for unconstrained optimization and nonlinear equations*, Society for Industrial Mathematics, 1996.
- [22] D.L. DONOHO, *Compressed sensing*, Information Theory, IEEE Transactions on, 52 (2006), pp. 1289–1306.
- [23] O. DORN AND D. LESSELIER, *Level set methods for inverse scattering*, Inverse Problems, 22 (2006), p. R67.
- [24] O. DORN, E.L. MILLER, AND C.M. RAPPAPORT, *A shape reconstruction method for EM tomography*, Inverse problems, 16 (2000), pp. 1119–1156.
- [25] M. ELWASEIF AND L. SLATER, *Quantifying tomb geometries in resistivity images using watershed algorithms*, Journal of Archaeological Science, 37 (2010), pp. 1424–1436.
- [26] M.A.T. FIGUEIREDO, R.D. NOWAK, AND S.J. WRIGHT, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, Selected Topics in Signal Processing, IEEE Journal of, 1 (2007), pp. 586–597.
- [27] A. GELAS, O. BERNARD, D. FRIBOULET, R. PROST, AND V. INSA, *Compactly supported radial basis functions based collocation method for level-set evolution in image segmentation*, IEEE Transactions on Image Processing, 16 (2007), pp. 1873–1887.
- [28] T. GOLDSTEIN AND S. OSHER, *The split bregman method for l_1 -regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.

- [29] D.H. GRIFFEL, *Applied functional analysis*, Dover publications, 2002.
- [30] J.H. HUBBELL AND S.M. SELTZER, *Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients 1 kev to 20 mev for elements $z=1$ to 92 and 48 additional substances of dosimetric interest*, tech. report, National Inst. of Standards and Technology-PL, Gaithersburg, MD (United States). Ionizing Radiation Div., 1995.
- [31] M. KASS, A. WITKIN, AND D. TERZOPOULOS, *Snakes: Active contour models*, International journal of computer vision, 1 (1988), pp. 321–331.
- [32] A. KIRSCH, *Characterization of the shape of a scattering obstacle using the spectral data of the far field operator*, Inverse problems, 14 (1998), pp. 1489–1512.
- [33] M. LUSTIG, D. DONOHO, AND J. PAULY, *Sparse mri: The application of compressed sensing for rapid mr imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195.
- [34] M.R. OSBORNE, B. PRESNELL, AND B.A. TURLACH, *A new approach to variable selection in least squares problems*, IMA journal of numerical analysis, 20 (2000), pp. 389–403.
- [35] —, *On the lasso and its dual*, Journal of Computational and Graphical statistics, 9 (2000), pp. 319–337.
- [36] S. OSHER AND R.P. FEDKIW, *Level set methods: an overview and some recent results*, Journal of Computational physics, 169 (2001), pp. 463–502.
- [37] —, *Level set methods and dynamic implicit surfaces*, Springer Verlag, 2003.
- [38] S. OSHER AND N. PARAGIOS, *Geometric level set methods in imaging, vision, and graphics*, Springer-Verlag New York Inc, 2003.
- [39] S. OSHER AND J.A. SETHIAN, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*, Journal of computational physics, 79 (1988), pp. 12–49.
- [40] N.R. PAL AND S.K. PAL, *A review on image segmentation techniques*, Pattern recognition, 26 (1993), pp. 1277–1294.
- [41] N. PARAGIOS AND R. DERICHE, *Geodesic active regions and level set methods for supervised texture segmentation*, International Journal of Computer Vision, 46 (2002), pp. 223–247.

- [42] D.L. PHAM, C. XU, AND J.L. PRINCE, *Current methods in medical image segmentation 1*, Annual review of biomedical engineering, 2 (2000), pp. 315–337.
- [43] J. ROMBERG, *Imaging via compressive sampling*, Signal Processing Magazine, IEEE, 25 (2008), pp. 14–20.
- [44] K. SAUER AND C. BOUMAN, *A local update strategy for iterative reconstruction from projections*, Signal Processing, IEEE Transactions on, 41 (1993), pp. 534–548.
- [45] R. TIBSHIRANI, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society. Series B (Methodological), (1996), pp. 267–288.
- [46] A. TSAI, A. YEZZI JR, AND A.S. WILLSKY, *Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification*, Image Processing, IEEE Transactions on, 10 (2001), pp. 1169–1186.
- [47] E. VAN DEN BERG AND M.P. FRIEDLANDER, *Probing the pareto frontier for basis pursuit solutions*, SIAM Journal on Scientific Computing, 31 (2008), pp. 890–912.
- [48] E. VAN DEN BERG AND M.P. FRIEDLANDER, *Sparse optimization with least-squares constraints*, SIAM Journal on Optimization, 21 (2011), pp. 1201–1229.
- [49] L.A. VESE AND T.F. CHAN, *A multiphase level set framework for image segmentation using the mumford and shah model*, International Journal of Computer Vision, 50 (2002), pp. 271–293.
- [50] L. VON AHN, M. BLUM, N. HOPPER, AND J. LANGFORD, *Captcha: Using hard ai problems for security*, Advances in CryptologyEUROCRYPT 2003, (2003), pp. 646–646.